

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: MICHAEL PAUL Examiner #: 73476 Date: 7/20/00
 Art Unit: 1646 Phone Number 305-7038 Serial Number: 08/656,811
 Mail Box and Bldg/Room Location: CM1/10E13 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Method for enhancing long term memory in a subject
 Inventors (please provide full names): Dusan Bratsch

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search SEQ IDNO: 1 CRFE
 Reverse translate SEQ IDNO: 1
 Interference and connected Database
 Thanks.

Christina Chan
 Edward Hart
 Technical Info Specialist
 STIC / Biotech
 CM1 12C14 Tel: 305-9203

STAFF USE ONLY

Searcher: _____

Searcher Phone #: _____

Searcher Location: _____

Date Searcher Picked Up: 7/25/00

Date Completed: 8/9/00

Searcher Prep & Review Time: _____

Clerical Prep Time: _____

Online Time: _____

Type of Search

NA Sequence (#) 1+1

AA Sequence (#) _____

Structure (#) _____

Bibliographic _____

Litigation _____

Fulltext _____

Patent Family _____

Other _____

Vendors and cost where applicable

Dialog _____

Questel/Orbit _____

Dr.Link _____

Lexis/Nexis _____

Sequence Systems 02

WWW/Internet _____

Other (specify) _____

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2000, 23:25:16 ; Search time 21.91 Seconds
(without alignments)
265.164 Million cell updates/sec

Title: US-08-656-811A-1

Perfect score: 1928

Sequence: 1 MCLDMSDFQIARWGLEM.....KYMKNLMDVCKAKGIOLKM 379

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/PTMUS_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	132	6.8	521	2	US-08-721-684C-2
2	132	6.8	521	2	US-09-005-870-2
3	128.5	6.7	266	2	US-08-319-866-8
4	124.5	6.5	1142	2	US-08-993-118-7
5	124.5	6.5	1142	3	US-08-845-528C-7
6	118.5	6.1	1363	1	US-08-425-061-23
7	118.5	6.1	1363	2	US-08-825-886-23
8	118.5	6.1	1852	1	US-08-425-061-24
9	118.5	6.1	1852	2	US-08-825-886-24
10	118.5	6.1	1863	1	US-08-425-061-16
11	118.5	6.1	1863	1	US-08-480-784-2
12	118.5	6.1	1863	1	US-08-483-553-2
13	118.5	6.1	1863	1	US-08-487-002-2
14	118.5	6.1	1863	1	US-08-483-554B-2
15	118.5	6.1	1863	1	US-08-798-691-4
16	118.5	6.1	1863	1	US-08-488-011B-2
17	118.5	6.1	1863	2	US-08-825-886-16
18	118.5	6.1	1863	2	US-08-603-753D-2
19	118.5	6.1	1863	3	US-08-825-487A-4
20	118.5	6.1	1863	4	PCT-US95-10202-2
21	118.5	6.1	1863	4	PCT-US95-10203-2
22	118.5	6.1	1863	4	PCT-US95-10220-2
23	113.5	5.9	1863	1	US-08-598-591-2
24	113.5	5.9	1863	1	US-08-798-691-2
25	113.5	5.9	1863	3	US-08-798-691-6
26	113.5	5.9	1863	3	US-08-825-487A-2
27	113.5	5.9	1863	3	US-08-825-487A-6
28	110.5	5.7	331	1	US-08-094-533B-10

29	110.5	5.7	331	1	US-08-276-860A-10	Sequence 10, Appl
30	110.5	5.7	331	1	US-08-444-393-10	Sequence 10, Appl
31	110.5	5.7	331	1	US-08-799-913-10	Sequence 10, Appl
32	110.5	5.7	331	2	US-08-711-893-10	Sequence 10, Appl
33	110.5	5.7	331	2	US-09-150-200-10	Sequence 10, Appl
34	110.5	5.7	331	2	US-09-150-201-10	Sequence 10, Appl
35	110.5	5.7	331	4	PCT-US95-15353-39	Sequence 39, Appl
36	109	5.7	1018	1	US-08-072-610-2	Sequence 2, Appl
37	109	5.7	1018	2	US-08-719-822B-2	Sequence 2, Appl
38	105	5.4	1202	2	US-08-425-061-22	Sequence 22, Appl
39	105	5.4	1202	2	US-08-825-886-22	Sequence 22, Appl
40	105	5.4	1706	2	US-08-459-568-2	Sequence 2, Appl
41	105	5.4	1706	2	US-08-399-411-2	Sequence 2, Appl
42	105	5.4	1706	3	US-08-516-859A-2	Sequence 2, Appl
43	104.5	5.4	1664	2	US-08-642-846-2	Sequence 2, Appl
44	104	5.4	331	4	PCT-US94-08119-10	Sequence 10, Appl
45	104	5.4	331	4	PCT-US94-08120-10	Sequence 10, Appl

ALIGNMENTS

```
RESULT 1
US-08-721-684C-2
; Sequence 2, Application US/08721684C
; Patent No. 5854016
; GENERAL INFORMATION:
; APPLICANT: Keegan, Kathleen S.
; TITLE OF INVENTION: No. 5854016el CREBA Isoform
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,684C
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/33487
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-721-684C-2

Query Match 6.8%; Score 132; DB 2; Length 521;
Best Local Similarity 21.5%; Pred. No. 0.0005;
Matches 91; Conservative 61; Mismatches 151; Indels 120; Gaps 22;

QY 19 EMPVYDQCGQCDKSTSRHGGDESLQPOGATLKLPFEDVIGAWMSSDLSGLD 78
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 8 EQSVLQMDRKLSL---SEPGETALMHTHFSEL-LDEFQNVLG-----Q 50

QY 79 ALGDNHERLHPFESNLLEFTSLTPDDSTVSKDILSLTLOPPTQPVNIPYASHGAEDFS 138
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 51 LLSD-----PFLS-----EKSESMVEVPS-----PTSPA--PLIQARHYSYLS 86
```

```

0Y 139 AETFEEN--HLSPDPS--PROV-----APINLEPETAASHMTVSPGGL 181
Db EEPHQSGFTHAASDSDEBVESEKKVLYSTERRPMTIKKEPI-----TEQPPGL- 188
0Y 182 GGMELASSLFFTELDPEPVNEN-----DSAVSIGGAEELIGS--PLSDVDVESTISFGP 234
Db 139 ----VPSTVLTITLIS--TFPEKESPLDMNAGGSSQOTLPIKTLRPHEDDFLINS-P 1922
0Y 235 SSPETSQSSIIETSSBELKYISTSIDASKRFPYSRSS-----KSGQSVKTS-D 284
Db 193 KEASVDQLHLRPTPPSSHSSDSEGLSPNRLHPFLSLQASHPVAMPGRPSALSTFPL 252
0Y 285 KAPKRTIR-----TPAOPVREPHVITM-----EILDKKDKKKLQKNNAIRYR 324
Db 253 TAPHLQSSGPLVLTREEKRTLVAEGYPIPKLPLTKSEKALIKIRKIKNISAOESR 312
0Y 335 MKKKGEAGIGIGEOBLEELNTKLKTIVDQD-----REIKYMNKL-----MEDQKAK 373
Db 313 RKKKRYMDSLEKVESCSSTENLRLRKVEYLEVLENTNRTLQDLQKLTVLWGKVSRRCKLA 3722
0Y 374 GIQ 376
Db 373 GIQ 375

```

```

RESULT 2
US-09-005-970-2
Sequence 2, Application US/09005970
Patent No. 5959079
GENERAL INFORMATION:
APPLICANT: - Keegan, Kathleen S.
TITLE OF INVENTION: No. 5959079e1 CREBA Isoform
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,970
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/721,684
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/33487
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-005-970-2

```

Query Match	6.8%;	Score 132;	DB 2;	Length 521;
Best Local Similarity	21.5%;	Pred. No. 0.0005;		
Matches 91;	Conservative 61;	Mismatches 151;	Indels 120;	Gaps 22

OY	19	EMPVOTGQGFCDLKSTSRHGGDESLDPOGATLKLEPEEDVIGAEHMESSDLSGFLD	78
Db	8	FOSVLQOMRKISEL-----SEGETALMHTHFSEL-LDEFSQNVLG-----Q	50
OY	79	ALGDNHERLAFPESSNLEFTSLTPRDSVSKDLSLSTLQRPQPVNIPRYASHGAEDFS	138
Db	51	LISD-----PFLS-----EKSESMEVEPS-----PTSPRA-PLIOAENHSYLS	86
OY	139	AETEFEN-----HLSPDS--PROV-----ABVINLREVELTASHMVTISPGCL	161
Db	87	EEPRKQSFPTAAISDSNDDEVESEKMYLSTEPSPATTIKERL-----TEBQPRGL-	138
OY	182	GGMELASESLFTLELDVFNFN-----DSAVSGSIGAEBELLS-PLSVDDVESTISFSGP	234
Db	139	----VPSTLTLTFAIS--TPFEKESRPLDMAAGDSSQOTLIPRKIKERHVEQDLNRS-P	132
OY	235	SSPEFSOSIIESSPELAKVYIYSTSIDASKRFSYSSS-----KSKQSVKTD-A	264
Db	193	KEASVDOLHLPRTPPSSHSDSEGLSPNRLAFPESSQASHPVAMPKRPASALSTSPRL	253
OY	285	KAPKTR-----TPAOPVPEHYVM-----EHLDKKDRKKLONKNAALRYR	324
Db	253	TAPKHLQSSGFLVLTBEERKRTLVAGFPIPKRLPLTSEBEKALKKIRKIKNKISADESR	312
OY	335	MKKKGGAOGIKGEEQLEELNTKIKUTYVDDI-----REIKYMKNL---MEDVCKAK	373
Db	313	RKKKEYMDSLEKKEVESCSTENLELRKKVEVELENTNRTLLDLOLRQLVLMGVKSVRTCKLA	372
OY	374	GIO 376	
Db	373	GTQ 375	

RESULT 3
US-08-319-866-8
Sequence 8, Application US/08319866
Patent No. 5929223
GENERAL INFORMATION:
APPLICANT: Tully, Timothy P.
APPLICANT: Yip, Jerry C.
APPLICANT: Regulski, Michael
TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF GENES
TITLE OF INVENTION: ASSOCIATED WITH LONG-TERM MEMORY
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millita Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
Zip: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/319,866
FILING DATE: 7-OCT-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: CSHL94-03
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 266 amino acids

REFERENCE/DOCKET NUMBER: LUD 5455
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1142
TYPE: amino acids
STRANDEDNESS: single-stranded
TOPOLOGY: linear
US-08-845-528C-7

Query Match 6.5%; Score 124.5; DB 3; Length 1142;
Best Local Similarity 22.0%; Pred. No. 0.0083;
Matches 82; Conservative 48; Mismatches 133; Indels 109; Gaps 17;
QY 5 LKSEDFQAREWGLEMPVVDGQFGDLKSTSRHGGDESLSLQPOGATLKLEPFEEDVLG 64
Db 616 LQEEFQSS-----LQSPVSTCS---SSTPSLSLQSPFPESQSPPEGPV--QSPLHSQSP 666
QY 65 AEWNESSDLGSFLDALGDNHERLHPFESNLEFTSLITPDDSTVSKDILSSTLQFP--- 120
Db 667 PEGMHSQSPLOSPEAPGEDSLSLQ-----IPOSPLGEDSLSS-LHFPQSP 715
QY 121 -----TOPVNIPIYASHGAEDF-----SAETEF-----ENHLSPPDPSPEQA 157
Db 716 EWEDLSPLHFPQFPQPG-EDFOSSLQSPVYICSSSTLSLQSPFPESQSPPEGPAQ-S 773
QY 158 PVNLPEVE-----ITASHMTVISPDGLLGMELASESLTTFELDFVFNDSAVG 207
Db 774 PL-QRPVSVSFYTLASLQSHESQSPPE----- 803
QY 208 SIGAEELLSPLS--VDDVESTISFGP--SSPETSQSSIIESSPELYKVIKSTSIDAS 263
Db 804 --GPAQSPLOSVPSPSPSTSSLSQSPSPSPSTSSLSKSSPSLQSPVISFSS 861
QY 264 KRFSPYSRSS-----KSKQVKTSDAKAPKRTTPAQPVPEHVIMEHLKDKKRLQ 315
Db 862 TSLSPSEESSPVDEVTSSDTLLESSTLDESLESEPLETYTLTDERVDELARLL- 920
QY 316 NKAAATRYRKK 327
Db 921 -----LKYQVKQ 927

RESULT 6
US-08-425-061-23
Sequence 23, Application US/08425061
Patent No. 5622829
GENERAL INFORMATION:
APPLICANT: KING, Mary-Claire
APPLICANT: FRIEDMAN, Lori
APPLICANT: OSTERMEYER, Beth
APPLICANT: ROWELL, Sarah
APPLICANT: LYNCH, Eric
APPLICANT: SZABO, Csilla
APPLICANT: LEE, Ming
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,061
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-061-23

Query Match 6.1%; Score 118.5; DB 1; Length 1363;
Best Local Similarity 19.2%; Pred. No. 0.039;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;
QY 13 AREWGLEMPVVDGQFGDLKST-----SRHGGDESLSLQ 47
Db 840 SRETSIEMSESLDAQI--LQNTFKVSKROSFPAPFSPGNAEECATFSAHG--SLKKQ 895
QY 48 POGATLKLEPFEEDVLGAEWNESSDLGSFLDALGDNHERLHPFESNLEFTSLITPDDST 107
Db 896 SPKVTFCEQKEEN-----QGNESNIKPVQT----- 922
QY 108 VSKDILSSTLQFP-----OPVNIPIYASHGAEDFSAETEF-----ENHLSPPD-----SP 153
Db 923 -----VNITAGFPVVGQKDPVDNAKSIKGGSRFLCSSQFRGNETGLITPNKHGLLQNP 977
QY 154 EQAPVNLPEVELTASHMTVI-----SPDGLLGMELASESLT-----FT 194
Db 978 YRTPPLFPPIKSFVTKCKKNLLEENFEHSMSPEREMGNENIPSTVTSIRNNIRENVFK 1037
QY 195 ELDFVFNFD--SAVGSIGGAELLSPLSDVDVESTISFS-GPS-----SPETSQ 241
Db 1038 EASSNINEVGSSTNEVGSINEIGS--SDENIQAEGLRNGRPKLNAMLRLGLVLQPEVYK 1095
QY 242 SSIIES--PELYK-----VISTSSIDASKRSPYSRSKSKQSVKTSDAKAPKRTTP 293
Db 1096 QSLPGSNCKHPKIKQIEYEVQTVNTD---FSPYLISDNLEQPMGSSHAS-----QV 1145
QY 294 AQPVPHEVIMEHLDKDKRKLQN---KNAAIRYRMKKKE-----AQGIKE 337
Db 1146 CSETPDLDLDDGELKEDTSPAENDIKESSAVFSKVQKGLSRLSPSPFTTHLAQGYRRG 1205
QY 338 EQEELNLTCLKTKVDDL 355
Db 1206 AKKLESEENLSSEDEL 1223

RESULT 7
US-08-825-886-23
Sequence 23, Application US/08825886
Patent No. 5821328
GENERAL INFORMATION:
APPLICANT: KING, Mary-Claire
APPLICANT: FRIEDMAN, Lori
APPLICANT: OSTERMEYER, Beth
APPLICANT: ROWELL, Sarah
APPLICANT: LYNCH, Eric
APPLICANT: SZABO, Csilla
APPLICANT: LEE, Ming
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
TITLE OF INVENTION: CANCER

NUMBER OF SEQUENCES: 24
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOIBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/825,886
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/425,061
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: OSMAN, Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 494-8700
 TELEFAX: (415) 494-8771
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 23:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1363 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-825-886-23

Query Match 6.1%; Score 118.5; DB 2; Length 1363;
Best Local Similarity 19.2%; Pred. NO. 0.039;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;

[illegible]

```

RESULT      8
US-08-425-061-24
: Sequence 24, Application us/08425061
: Patent No. 5622829
: GENERAL INFORMATION:
: APPLICANT: KING, Mary-Claire
: APPLICANT: FRIEDMAN, Lori
: APPLICANT: OSTENMEYER, Beth
: APPLICANT: ROWELL, Sarah
: APPLICANT: LYNCH, Eric
: APPLICANT: SZABO, Csilla
: APPLICANT: LEE, Ming
: TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
: TITLE OF INVENTION: CANCER
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: City: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/425.061
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, Richard A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 494-8700
: TELEFAX: (415) 494-8771
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 24:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1852 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-425-061-24

```

Query Match	6.1%; Score 118.5; DB 1;	Length 1852;
Best Local Similarity	19.2%;	Pred. No. 0.063;
Matches 84; Conservative	70; Mismatches 135;	Indels 149; Gaps 20;
QY	13 AREGLEMPVQTGGQFGDKST----	-----SRHGDESLSQ 47 :
Dd	840 SRETSIEMESELDAQY--LQNTFKVKRQSFAPSPGNAEECATSAHSG--SLAKKQ 895	: :
QY	48 PQGATLKLEPFEEDVLGAENMESSDLGSFLDALGDNHERLHPFESNLLEFTSLITPD DST 107	: :
Dd	896 SPKVTFECEQKEEN-----	---QCKNESNIKPVQT----- 922
QY	108 VSKDILSSTLOFFT-----QPVNIPLYASHGAEDFSAEATER-----ENHLSPDD-----SP 153	: :
Dd	923 -----VNITAGFPVVGOKDFVDNAKGSIKGGSRFCSSQPRGNETGLITPNKHGLLQNP 977	: :
QY	154 EQAVPINLEPVELTASHMTVI-----SPDGLLGMLASESLT-----FT 194	: :
Dd	978 YRIPLPIKSVFYTKCKKNLEENFEHSMSPREGMENIPVTVISRNNTRENVEFK 1037	: :
QY	195 ELDFEVPND--SAYGSIGGABELGLPSLVDDVVSTISFS-GPS-----SPETSQ 241	: :

;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-425-061-16

Query Match 6.1%; Score 118.5; DB 1; Length 1863;
Best Local Similarity 19.2%; Pred. No. 0.064;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVQTDGQFDLKST-----SRHGGDESLSQ 47
DB 840 SRETSIEMESELDAQY--LQNTFKVSKRSQSFAPFSPNPNABECATFSAHSG--SLKKQ 895
QY 48 PQGATLKLEPFEDVLGAENWESSDLGSLDALGDNHERLHPFESNLLEFTSLITPDDST 107
DB 896 SPKVTFECEQKEEN-----QCKNESNIKPVQT----- 922
QY 108 VSKDILSSTLOFT-----QPVNIPLYASHGAEDFSAETEF---ENHLSPPD-----SP 153
DB 923 -----VNITAGFPVVGQKDPVDNAKCSIKGSRFLSSQFRGNETGLITPNKHGLLQNP 977
QY 154 EQVAPVINLEPVELTASHMTVI-----SPDGLLGMELASESLT-----FT 194
DB 978 YRIPPLPIKSFVTKCKKNLLEENFEHSMSPEREMGNENIPSTVTSIRNNIRENVFK 1037
QY 195 ELDFVNFND--SAVSGSIGGAELGLSPLSDVDVESTISFS--GPS-----SPETSQ 241
DB 1038 EASSNINEVGSSTNEVGSSINEIGS--SDENIQAEIAGRNRPKLNAMLRLGVLPQPEYK 1095
QY 242 SSIIESS---PELYK-----VISTSSIDASKRFSPPYSRKSQSVKTSDAKAPRKTRTP 293
DB 1096 QSLPGSNCKHPKQKQEEYEVQVTNTD-----FSPYLISDNLEQPMGSSHAS-----QV 1145
QY 294 AQPVEHVIMEHLDDKDRKKLQN---KNAATRYRMKKGE-----AQGIKGE 337
DB 1146 CSETPDLDLDDGEIKEDTFAENDIKESSAVFSVKQKGLSRSPSPFTTHLAQGYRRG 1205
QY 338 EQELEELNTKLTKRVDDL 355
DB 1206 AKLESSEENLSSEDEEL 1223

RESULT 11

US-08-480-784-2
; Sequence 2, Application US/08480784
; Patent No. 5693473

GENERAL INFORMATION:

; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

;
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,784
; FILING DATE:
; CLASSIFICATION: 435

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300

;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid

;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-784-2

Query Match 6.1%; Score 118.5; DB 1; Length 1863;

Best Local Similarity 19.2%; Pred. No. 0.064;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVQTDGQFDLKST-----SRHGGDESLSQ 47
DB 840 SRETSIEMESELDAQY--LQNTFKVSKRSQSFAPFSPNPNABECATFSAHSG--SLKKQ 895
QY 48 PQGATLKLEPFEDVLGAENWESSDLGSLDALGDNHERLHPFESNLLEFTSLITPDDST 107
DB 896 SPKVTFECEQKEEN-----QCKNESNIKPVQT----- 922
QY 108 VSKDILSSTLOFT-----QPVNIPLYASHGAEDFSAETEF---ENHLSPPD-----SP 153
DB 923 -----VNITAGFPVVGQKDPVDNAKCSIKGSRFLSSQFRGNETGLITPNKHGLLQNP 977
QY 154 EQVAPVINLEPVELTASHMTVI-----SPDGLLGMELASESLT-----FT 194
DB 978 YRIPPLPIKSFVTKCKKNLLEENFEHSMSPEREMGNENIPSTVTSIRNNIRENVFK 1037
QY 195 ELDFVNFND--SAVSGSIGGAELGLSPLSDVDVESTISFS--GPS-----SPETSQ 241
DB 1038 EASSNINEVGSSTNEVGSSINEIGS--SDENIQAEIAGRNRPKLNAMLRLGVLPQPEYK 1095
QY 242 SSIIESS---PELYK-----VISTSSIDASKRFSPPYSRKSQSVKTSDAKAPRKTRTP 293
DB 1096 QSLPGSNCKHPKQKQEEYEVQVTNTD-----FSPYLISDNLEQPMGSSHAS-----QV 1145
QY 294 AQPVEHVIMEHLDDKDRKKLQN---KNAATRYRMKKGE-----AQGIKGE 337
DB 1146 CSETPDLDLDDGEIKEDTFAENDIKESSAVFSVKQKGLSRSPSPFTTHLAQGYRRG 1205
QY 338 EQELEELNTKLTKRVDDL 355
DB 1206 AKLESSEENLSSEDEEL 1223


```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Innen, Jeffrey L.
; REGISTRATION NUMBER: 28,
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-002-2

```

Query Match	6.1%	Score 118.5;	DB 1;	Length 1863;
Best Local Similarity	19.2%;	Pred. No. 0.064;		
Matches	84;	Conservative	70;	Mismatches 135; Indels 149; Gaps 20;
QY	13	AREWGLEMPVVQDQGFCDLKST-----	SRUGGDESLSLQ	47
DB	840	SRETSIEMEESLDAQY--LQNTFKVSKROSFAPFSPGNAEECATFSAHSG--SLKKQ	895	
QY	48	PQAGATKLEFEEDVLCAEWMESSDLGSFLDAGDNHERLHPFESNLEFFSLITPD DST	107	
DB	896	SPKVTFCEQKEEN-----	QKQNESNIKPVGT-----	922
QY	108	VSKDILSSTLQFPT-----	QPVNIPLVASHGAEDFSAETEF---ENHLSPPD-----	SP 153
DB	923	-----VNITAGFPVVGQKDPVDNAKCSIKGGSFCLSSQPRGNETGLITPNKHGLQNP	977	
QY	154	EQVAPVINLPVELTAGHMTVI-----	SPDGLLGGMELASESLT-----	FT 194
DB	978	YRTPPLPPIKSEVKTCKKNLEENFEHNSMSPERENGNIPTSTVISRNNIRENVFK	1037	
QY	195	ELDFEVND--SAGVSTGGAEELGSPLSVDVESTISFS--GPS-----	SPETSQ	241
DB	1038	EASSNNINEVSGSTNEVSGSTNEIGTS--SDENIOAELGRNRPKLNAMLRLGVQLQPEYVK	1095	
QY	242	SSIIESS--PELYK-----	VISPTSIDASKRFSPYSRSSKSQVKTSDAKAPKTRTP	293
DB	1096	QSLPG\$CKKHPEIKQBYEEVVQVNTD---FSPYIISDNLEQPMGSSHAS-----	QV 1145	
QY	294	AQPVPHVIMHEHLDKKDRKKLQN---KNAAIRRMKKKGE-----	AQKIGE	337
DB	1146	CSETPDLLDGEIKEDTSAENDIKESSAVFSKVQKGLSRSPSPFTHHLAQGYRRG	1205	
QY	338	EQELEELNLTKLTKVDOL	355	
DB	1206	AKKLESSEENLSDDEEL	1223	

```

RESULT 14
US-08-483-554B-2
; Sequence 2, Application US/08483554B
; Patent No. 5747282
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked B
; TITLE OF INVENTION: Susceptibili
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, H

```

```

, STREET: 1201 New York Avenue, N.W., Suite 1000
, CITY: Washington
, STATE: DC
, COUNTRY: USA
, ZIP: 20005
, COMPUTER READABLE FORM:
, MEDIUM TYPE: Floppy disk
, COMPUTER: IBM PC compatible
, OPERATING SYSTEM: PC-DOS/MS-DOS
, SOFTWARE: PatentIn Release #1.0, Version #1.30
, CURRENT APPLICATION DATA:
, APPLICATION NUMBER: US/08/483,554B
, FILING DATE: 07-JUN-1995
, CLASSIFICATION: 514
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 08/409,305
, FILING DATE: 24-MAR-1995
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 08/348,824
, FILING DATE: 29-NOV-1994
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 08/308,104
, FILING DATE: 16-SEP-1994
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 08/300,266
, FILING DATE: 02-SEP-1994
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 08/289,221
, FILING DATE: 12-AUG-1994
, ATTORNEY/AGENT INFORMATION:
, NAME: Ihnen, Jeffrey L.
, REGISTRATION NUMBER: 28,957
, REFERENCE/DOCKET NUMBER: 24884-109347
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 202-962-4810
, TELEFAX: 202-962-8300
, INFORMATION FOR SEQ ID NO: 2:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 1863 amino acids
, TYPE: amino acid
, TOPOLOGY: linear
, MOLECULE TYPE: protein
US-08-483-554B-2

```

```

Query Match      6.1% Score 118.5; DB 1; Length 1863;
Best Local Similarity 19.2%; Pred. No. 0.064;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;
```

Qy	13	ARENGLEMPVVDTGQFGDLKST-----SRUGGESLSLQ	47
Db	840	SRETSIEMERSELDQAQY--LONTFKVKROSFAPFSPNPGNAEECATFSAHSG--SLKKQ	895
Qy	48	POGATLKLEPFEEDVLGAENWESSDGLSFDALGDNIHERLHPFESNLLEFTSLITPD DST	107
Db	896	SPKYTFECEQKEEN-----QCKNESNIKPVOT-----	922
Qy	108	VSKDILSTTQFPPT-----QPVNPLVASHGAEDFSAETEF---ENHLSPDD-----SP	153
Db	923	-----VNITAGFPVVGOKDPVDNAKGSIKGGSRFLCSSQFRGNETGLITPNKHGLQN P	977
Qy	154	EOVAPVINLEPVELTASHMTVI-----SPDGLGGMELASEST-----FT	194
Db	978	YRIPLPPIKSVTKCKKNLLLENFEEHSMSPREMGENEIPVTVISRNNTRENVPK	1037
Qy	195	ELDVPNFND--SAYGSGIGABELGLPSLVDDVESTISFS-GPS-----SPETSQ	241
Db	1038	EASSNNINEVGSSNTEVGSSINEIGS--SDENIQAEGLGRNGPKLNAMLRLGVQLQPEYK	1095
Qy	242	SSIIESS---PELYK-----VISTSSIDAKRFSPYSRSSKSOSKVTSIDAKAPKRTRTP	293
Db	1096	QSLFGSNCKHPEIKKKQEYEVQVNTD-----FSPYLISDNLEQPMGSSHAS-----QV	1145

RESULT 14
US-08-483-554B-2
: Sequence 2, Application US/08483554B
: Patent No. 5747282
: GENERAL INFORMATION:
: APPLICANT: Skolnick, Mark H.
: APPLICANT: Goldgar, David E.
: APPLICANT: Miki, Yoshio
: APPLICANT: Swenson, Jeff
: APPLICANT: Kamb, Alexander
: APPLICANT: Harshman, Keith D.
: APPLICANT: Shattuck-Eidens, Donna M.
: APPLICANT: Tavtigian, Sean V.
: APPLICANT: Wiseman, Roger W.
: APPLICANT: Futreal, P. Andrew
: TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
: TITLE OF INVENTION: Susceptibility Gene
: NUMBER OF SEQUENCES: 85
: ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP

QY 294 AQPVEHVIMEHLDKDKRKKLQN---KNAATRYRMKKKGE-----AQGIKGE 337
Db 1146 CSETPDDLLDDGETKEDTSAENDIKESSAVFSKSVQKGLSRSPSPFTHTHLAQGYRRG 1205
QY 338 EQELEELNTKLTKVDDL 355
Db 1206 AKLESSEENLSEDEEL 1223

RESULT 15
US-08-798-691-4
; Sequence 4, Application US/08798691
; Patent No. 5750400
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRCAL Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONCORMED
; STREET: 200 Perry Parkway
; CITY: Gaithersburg
; STATE: MD
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,691
; FILING DATE: 12-Feb-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas Gallegos
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: PA-0054CIP
; TELEPHONE: 301-527-2051
; TELEFAX: 301-208-6997
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCAL
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
US-08-798-691-4

Query Match 6.1%; Score 118.5; DB 1; Length 1863;
Best Local Similarity 19.2%; Pred. No. 0.064;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVYVQDGFQGLKST-----SRHGGDESLSLQ 47
Db 840 SRETSIEMESELDAQY--LQNTFKVSKRQSFALFNPNGNAEECATFSAHSG--SLKKQ 895
QY 48 PQGATLKLEPFEEEDVLGAENWESSDLGSLDALGDNHERLHPFESNULLEFTSLITPDST 107

Db 896 SPKVTFECEQKEEN-----QGNESNIKPVQT----- 922
QY 108 VSKDILSSTLQFPT-----QPVNIPLYASHGAEDFSAETFE---ENHLSPPD-----SP 153
Db 923 -----VNITAGFPVVGQDKDPVNAKCSIKGSRFCFLSSQFRGNETGLITPNKHGLLQNP 977
QY 154 EQAPVINLEPVELTASHMTVI-----SPDGLLGGMELASESLT-----FT 194
Db 978 YRIPPLPIKSFVKTKCKKNLLEENFEHSMSPEREMGNENIPSTVSTISRNINRENVFK 1037
QY 195 ELDFVNEND--SAVGSTGGABELLGSLVDVDDVESTISFS-GPS-----SPETSQ 241
Db 1038 EASSNINEVGSSTNEVGSSINEIGS--SDENIQAEGLRNRGPKINAMLRGLGVLPQPEYK 1095
QY 242 SSIIESS---PELYK-----VISTSSIDASKRFSPPYSSKSKSQSVKTSDAKAPRKTRTP 293
Db 1096 QSLPGSNCKHPKIKKQEEYEVVQTVNTD---FSPYLIISDNLEQPMGSSHAS-----QV 1145
QY 294 AQPVEHVIMEHLDKDKRKKLQN---KNAATRYRMKKKGE-----AQGIKGE 337
Db 1146 CSETPDDLLDDGETKEDTSAENDIKESSAVFSKSVQKGLSRSPSPFTHTHLAQGYRRG 1205
QY 338 EQELEELNTKLTKVDDL 355
Db 1206 AKLESSEENLSEDEEL 1223

Search completed: August 8, 2000, 02:39:09
Job time: 11633 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2000, 23:19:16 ; Search time 25.31 Seconds
(without alignments)
354.683 Million cell updates/sec

Title: US-08-656-811A-1

Perfect score: 1928

Sequence: 1 MELDLWSEDFQLAREWGLEM.....KYNKNLMEDVCRAKGIQLKM 379

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1923	99.7	378	1 W41508	Aplysia CAMP-respo
2	132	6.8	521	1 W37934	Mouse CAMP regulat
3	128.5	6.7	266	1 R91295	Drosophila dCREB1
4	124.5	6.5	1142	1 R81346	Tumour rejection a
5	121.5	6.3	1863	1 R81540	BRCAL mutant from
6	120.5	6.2	654	1 R13950	E2Aalpha protein.
7	120.5	6.2	1863	1 R81514	BRCAL mutant from
8	120.5	6.2	1863	1 R81543	BRCAL mutant from
9	119.5	6.2	1503	1 W48845	Human receptor tyr
10	118.5	6.1	1261	1 R81523	BRCAL mutant from
11	118.5	6.1	1363	1 R81525	BRCAL mutant from
12	118.5	6.1	1363	1 R99447	BRCAL allele #582
13	118.5	6.1	1363	1 W10010	Protein encoded by
14	118.5	6.1	1363	1 W79896	Tumourigenic BRCAL
15	118.5	6.1	1442	1 R81526	BRCAL mutant from
16	118.5	6.1	1540	1 R81483	BRCAL mutant from
17	118.5	6.1	1540	1 R81527	BRCAL mutant from
18	118.5	6.1	1619	1 R81528	BRCAL mutant from
19	118.5	6.1	1669	1 R81530	BRCAL mutant from
20	118.5	6.1	1828	1 R81531	BRCAL mutant from
21	118.5	6.1	1852	1 R99448	BRCAL allele #77 t
22	118.5	6.1	1852	1 W10011	Protein encoded by
23	118.5	6.1	1852	1 W79897	Tumourigenic BRCAL
24	118.5	6.1	1863	1 R76641	BRCAL protein. Det
25	118.5	6.1	1863	1 R91208	BRCAL, breast and
26	118.5	6.1	1863	1 R81481	BRCAL, New nucleic
27	118.5	6.1	1863	1 R81485	BRCAL mutant from
28	118.5	6.1	1863	1 R81486	BRCAL mutant from
29	118.5	6.1	1863	1 R81490	BRCAL mutant from
30	118.5	6.1	1863	1 R81497	BRCAL mutant from
31	118.5	6.1	1863	1 R81500	BRCAL mutant from
32	118.5	6.1	1863	1 R81524	BRCAL mutant from
33	118.5	6.1	1863	1 R81529	BRCAL mutant from

ALIGNMENTS

RESULT 1

W41508
ID W41508 standard; Protein; 378 AA.

AC W41508;
DT 22-JUN-1998 (first entry)
DE Aplysia CAMP-response element binding protein 2.
KW APCR-2: CAMP-response element binding protein-2; snail;
KW memory loss; Alzheimer's disease; amnesia; ischaemia; head trauma;
KW neuronal injury; Parkinson's disease; senility; therapy.
OS Aplysia californica.
FH Key Location/Qualifiers
FT Region 73..108
FT /note= "leucine heptad repeat region"
FT Modified_site 150..153
FT /note= "MAP kinase phosphorylation site"
FT Modified_site 234..237
FT /note= "MAP kinase phosphorylation site"
FT Modified_site 271..174
FT /label= "protein kinase C phosphorylation site"
FT Domain 306..378
FT /note= "basic region/leucine zipper domain,
FT interacts with ApC/EBP"

WO9746257-A1.

11-DEC-1997.

03-JUN-1997; U09438.

03-JUN-1996; US-656811.

(UYCO) UNIV COLUMBIA NEW YORK.

Bartsch D, Ghirardi M, Kandel ER;

WPI: 98-051903/05.

N-PSDB: V04079.

Enhancing long-term memory in subjects whose cAMP-responsive gene is repressed - used to treat long-term memory defects, e.g. age-related memory loss, Alzheimer's disease

Example 2; Page 73-74; 100pp; English.

This protein comprises the CAMP-response element binding protein-2, i.e. APCR-2, of the marine snail Aplysia. APCR-2 is a

transcription factor and repressor of long-term facilitation in Aplysia neurons. It is constitutively expressed in sensory

neurons. APCR-2 is a homologue of human CREB-2 and mouse ATF-4.

The amino acid sequence of APCR-2 was deduced from a polynucleotide

sequence (see V04079) obtained from 2 clones isolated from an Aplysia

central nervous system cDNA library. The invention provides a method

of enhancing long-term memory in a subject whose cAMP-responsive gene

expression is repressed due to binding of a CREB-2 to a protein or

DNA associated with CAMP-responsive gene expression, or both. The

method involves administering to the subject a compound capable of

interfering with such binding so as to derepress CAMP-responsive gene

expression and thereby enhance long-term memory. Such compounds

include anti-CREB-2 antibodies or a compound capable of altering

phosphorylation of CREB-2. The method is used to treat e.g.

age-related memory loss, Alzheimer's disease, amnesia, ischaemia,

shock, head trauma, neuronal injury, toxicity or degradation,

CC Parkinson's disease or senility (claimed).

Sequence 378 AA;

50

34 118.5 6.1 1863 1 R81532 BRCAL mutant from
35 118.5 6.1 1863 1 R81533 BRCAL mutant from
36 118.5 6.1 1863 1 R81534 BRCAL mutant from
37 118.5 6.1 1863 1 R81535 BRCAL mutant from
38 118.5 6.1 1863 1 R81505 BRCAL mutant from
39 118.5 6.1 1863 1 R81509 BRCAL mutant from
40 118.5 6.1 1863 1 R81537 BRCAL mutant from
41 118.5 6.1 1863 1 R81538 BRCAL mutant from
42 118.5 6.1 1863 1 R81542 BRCAL mutant from
43 118.5 6.1 1863 1 R81544 BRCAL mutant from
44 118.5 6.1 1863 1 R81545 BRCAL mutant from
45 118.5 6.1 1863 1 R81546 BRCAL mutant from

Query Match 99.7%; Score 1923; DB 1; Length 378;
 Best Local Similarity 100.0%; Pred. No. 8.6e-150;
 Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELDLASEDQALREWGLENVVTQDQFGDLKSTSRHGDESLSLQPOGATLKLEPFE 60
 DB 1 MELDLASEDQALREWGLENVVTQDQFGDLKSTSRHGDESLSLQPOGATLKLEPFE 60
 QY 61 DVLGAEMWESSDGLSFDALGDHNERLHPFESNLEFTSLITPDSTVSKDILSSTLQFP 120
 DB 61 DVLGAEMWESSDGLSFDALGDHNERLHPFESNLEFTSLITPDSTVSKDILSSTLQFP 120
 QY 121 TQPVNIPLYASHGAEDFSAETEFENHLSPPDSPEQVAPVINLEPVELTASHMTVISPDGL 180
 DB 121 TQPVNIPLYASHGAEDFSAETEFENHLSPPDSPEQVAPVINLEPVELTASHMTVISPDGL 180
 QY 181 LGMELASESLTTELDVFNFDNSAVSGISGABELLGSPSLVDVESTISFGSPSPETS 240
 DB 181 LGMELASESLTTELDVFNFDNSAVSGISGABELLGSPSLVDVESTISFGSPSPETS 240
 QY 241 QSSIIESSPELYKVISSIDASKRFSPYSRSKSKSVKTSDAKAPRKTRTPAQVPVEH 300
 DB 241 QSSIIESSPELYKVISSIDASKRFSPYSRSKSKSVKTSDAKAPRKTRTPAQVPVEH 300
 QY 301 VIMEHLDDKDRKKLQNKNAARIYRMKKKGAEQIKGEELEELNTKLTKVDDLOREIK 360
 DB 301 VIMEHLDDKDRKKLQNKNAARIYRMKKKGAEQIKGEELEELNTKLTKVDDLOREIK 360
 QY 361 YMKNLMDVCKKAGIQLK 378
 DB 361 YMKNLMDVCKKAGIQLK 378

RESULT 2
 W37934
 ID W37934 standard; Protein; 521 AA.
 AC W37934;
 DT 10-AUG-1998 (first entry)
 DE Mouse CAMP regulatory element binding protein.
 KW Mouse CAMP regulatory element binding; mCREBa; antibody; inhibition;
 KW mCREBa modulator.
 OS Mus sp.
 PN W09812910-A2.
 PD 02-APR-1998.
 PF 26-SEP-1997; U17288.
 PR 27-SEP-1996; US-721684.
 PA (ICOS-) ICOS CORP.
 PI Keegan KS;
 DR WPI; 98-230326/20.
 DR N-PSDB; V29204.
 PT Murine CAMP regulatory element binding protein A, mCREBa - useful
 PT for systematic analysis of structure and function of mCREBa and
 PT identification of those molecules with which it will react
 PS Claim 1; Pages 30-31; 41pp; English.
 CC This is the amino acid sequence of the mouse CAMP regulatory element
 CC binding (mCREBa) protein. The products of the mCREBa such as
 CC antibodies can be used for the recombinant production of the protein,
 CC to identify novel genes encoding binding partner polypeptides for
 CC mCREBa, for the systematic analysis of the structure and function of
 CC mCREBa and identification of those molecules with which it will react,
 CC to identify inhibitors of mCREBa binding to other natural binding
 CC partners and to generate rodents that fail to express a functional
 CC mCREBa or express a variant mCREBa, useful as models for studying the
 CC activities of mCREBa and mCREBa modulators in vivo.
 SQ Sequence 521 AA;

Query Match 6.8%; Score 132; DB 1; Length 521;
 Best Local Similarity 21.5%; Pred. No. 0.0074;
 Matches 91; Conservative 61; Mismatches 151; Indels 120; Gaps 22;

QY 19 EMPVVTQDQFGDLKSTSRHGDESLSLQPOGATLKLEPFEEDVLGAEMWESSDGLSFLD 78
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

DB 8 EQSVLOWDRKLSL---SEPCETEALMYHTHFSEL-LDEFSONVLG-----Q 50
 QY 79 ALGDNHERLHPFESNLEFTSLITPDSTVSKDILSSTLQPTQPVNIPLYASHGAEDFS 138
 DB 51 LLSD-----PFLS-----EKSESMEVEPS-----PTSPA--PLIQAEHSYLS 86
 QY 139 AETEFEN---HLSPDPS--PEQV-----ADVINLEPVELTASHMTVISPDGLL 181
 DB 87 EEPRTQSPFTHAATSDSFENDEEVESEKWLSTEPSATIKKEPI-----TEQPPGL- 138
 QY 182 GGMELASESLTTELDVFNFN-----DSAVSGISGABELLGSPSLVDVESTISFGSP 234
 DB 139 -----VPVTLTITAIIS--TPFEKEESPLDMNAGDSSCOTLIPKIKLPHEVDQFLNFS-P 192
 QY 235 SSPETSOSIIESSPELYKVISSIDASKRFSPYSRSS-----KSKQSVKTS-D-A 284
 DB 193 KEASVDQLHLPTPPSSHSSDSSEGLSPNRLHPFLSQAHSVPVRAMPGRPSALUSTSPLL 252
 QY 285 KAPRKTR-----TPAQVPVEHVM-----EHLDDKDRKKLQNKNAARIYR 324
 DB 253 TAPHKLOGSGPLVLTTEEKRTLVAEGYPIPTKLPLTKSEKALKIRRKIKNKISAQESR 312
 QY 325 MKKKGAEQIKGEELEELNTKLTKVDDLO-----REIKYMKNL-----MEDVCKAK 373
 DB 313 RKKKEYMDSLEKKVESCSTENLELRKKVEVDENTNRTLQLOQLQTLVMGKVSRTCKLA 372
 QY 374 GIQ 376
 DB 373 GTQ 375

RESULT 3
 R91295
 ID R91295 standard; Protein; 266 AA.
 AC R91295;
 DT 09-JUL-1996 (first entry)
 DE Drosophila dCREB1 protein.
 KW CAMP-responsive transcriptional activator; dCREB1;
 KW long-term memory; transgenic animal; Insect.
 OS Drosophila sp.
 FH Key Location/Qualifiers
 FT domain 1..215
 FT /label= Activation_domain
 FT region 216..237
 FT /label= Basic_region
 FT /note= "contains a high proportion of Arg and Lys
 FT residues"
 FT domain 238..264
 FT /label= Leucine_zipper
 FT /note= "contains periodic Leu residues"
 PN W09611270-A1.
 PD 18-APR-1996.
 PF 06-OCT-1995; U13198.
 PR 07-OCT-1994; US-319866.
 PR 21-DEC-1994; US-361063.
 PA (COLD-) COLD SPRING HARBOR LAB.
 PI Regulski M, Tully TP, Yin JC;
 DR WPI; 96-209851/21.
 DR N-PSDB; T14025.
 PT Novel cyclic 3',5'-adenosine monophosphate responsive
 PT transcriptional activator gene - used in stimulation and enhancement
 PT of longterm memory
 PS Claim 32; Page 100-101; 160pp; English.
 CC The Drosophila CREB1 protein (R91295) is an enhancer-specific
 CC activator and a member of the CREB/ATF family. It is the product
 CC of a cDNA clone (T14025) obt'd. from a Drosophila head cDNA
 CC library. dCREB1 mediated transcriptional activation from
 CC CAMP-responsive element-contg. reporters in the Drosophila L2 cell
 CC line, but this activation was not dependent of protein kinase
 CC activity. Expression of the dCREB1 gene is associated with long-term
 CC memory in Drosophila. dCREB1 can be used to manipulate memory, and
 CC transgenic insects expressing dCREB1 can be used to assess the effects
 CC of drugs on long-term memory formation.

SQ	Sequence	266 AA;
----	----------	---------

Query Match 6.7%; Score 128.5; DB 1; Length 266;
Best Local Similarity 24.5%; Pred. No. 0.0053;
Matches 76; Conservative 50; Mismatches 99; Indels 85

QY	91	ESNLLETSII	TPDD	TSVK-----DILS	STLOFT	QTPQ	VNI	PIYASHGA	134
		:	:	:	:	:	:	:	:
Db	5	ENMFST	FTSL---	DAATATT	GEFLM	NSPQ	EAGD	MLDLS	DFNIMGNL-----A
		:	:	:	:	:	:	:	:
QY	135	EDRSAET	EPEFNH	LSPP	DSPEQ	VAP-VIN	LEPVEL	TASHMT	VISPDGLLG
		:	:	:	:	:	:	:	:
Db	55	DDFQTS	-----	ASPASE	DKMTP	FFVYD	TNVF	SVFKN---	TEDTLIG--DIDNVG
		:	:	:	:	:	:	:	:
QY	194	TEL-DFV	NENDS	AVGS	IGGA	ELLG	SPL---SV	DDVEST	ISFSPSP
		:	:	:	:	:	:	:	:
Db	102	TELKMF	DLVD	SEINN	-----	GTP	KOEK	DDLEF	-----SRSQ
		:	:	:	:	:	:	:	:
QY	250	ELK	VIST	SS	ID	ASKR-F	SPYS	RSSK---SK	QSVKTS
		:	:	:	:	:	:	:	:
Db	145	---	KST	SAS	PAD	AA	ACAS	FSS	SSCKSYSA
		:	:	:	:	:	:	:	:
QY	297	VPEH	VIMH	LK	DKR	KLON	KNAAIR	YMRK	KGEAQ
		:	:	:	:	:	:	:	:
Db	202	LPP-VIP	KGQD	VAS	MKEAR	TEA	RRSR	KMERM	-----SQ
		:	:	:	:	:	:	:	:
QY	357	REIK	MKNLM	366					
		:	:	:	:	:	:	:	:
Db	254	AOV	ALK	KLK	1263				
		:	:	:	:	:	:	:	:

RESULT

ID W01546 standard; Protein; 1142 AA.
 AC W81546;
 DT 01-MAR-1999 (first entry)
 DE Tumour rejection antigen precursor MAGE-C1.
 KW MAGE-C1; human; tumour rejection antigen precursor; TRAP;
 KW therapy; diagnosis.
 OS Homo sapiens.
 PN W09849184-AL.
 PS 05-NOV-1998.
 PD 24-APR-1998; U08493.
 PF 25-APR-1997; US-845528.
 PR (LUDW-) LUDWIG INST CANCER RES.
 PA Boon-Fallieur T, De Smet C, Lucas S;
 PI WPI: 99-024041/02.
 DR N-PSDB: V69720.
 DR Tumour rejection antigen precursors - used for determining presence
 PT of cytolytic T cells specific for complexes of a human leukocyte
 PT antigen
 PT Disclosure; Page 43-46; 84pp; English.
 PS This is the amino acid sequence of novel human tumour rejection
 PS antigen precursor (TRAP) MAGE-C1, a novel member of the MAGE
 PS family that may be recognised by cytotoxic T cells, leading to
 PS lysis of the tumour cells which express it. MACE-C1 and MACE-C2
 PS (see W81547) are expressed in a variety of tumours and in normal
 PS testis cells, but not by other normal cells. MAGE-C1 cDNA (see
 PS V69720) was isolated from a melanoma L6373-MEL cDNA. MACE-C1 and
 PS MACE-C2 cDNAs (see V69726) are claimed, as are: expression vectors;
 PS transformed or transfected cell lines (e.g. COS and CHO); an
 PS isolated TRAP encoded by the cDNAs; a kit useful in a PCR based
 PS assay; a method for determining expression of a MACE-C1 gene using
 PS the kit; a polypeptide comprising a number of tumour rejection
 PS antigens derived from MAGE-C1 or MAGE-C2; and a polypeptide comprising
 PS at least one tumour rejection antigen derived from MAGE-C1 or
 PS MAGE-C2 and at least one other tumour rejection antigen. MACE-C1
 PS and MACE-C2 can be used in a method for determining the presence of
 PS cytolytic T cells specific for complexes of a human leukocyte
 PS antigen (HLA).
 PS Sequence 1142 AA;
 SQ

Query Match

Query Match 6.5%; Score 124.5; DB 1; Length 1142;
Best Local Similarity 22.0%; Pred. No. 0.099;
Matches 82; Conservative 48; Mismatches 133; Indels 109;

QY	5	LWSEDFOLAREWGLEMPVOTDQGFGLDKTSRHGGDESLSLOPQATUKLEPFEEDVLG	64
Db	616	LQGEFOSS-----LQSPVSIC-----SSTPSLQSPFESSQSPGPPV--QSPHLSPQSP	666
QY	65	AENWESSDLGSFDLALGDHNERLHPRESNLEFTSLTPDDSVSKDILSLSLTQFP----	120
Db	667	PEGMHSOSPQSPESAPEGDSLSPLQ-----IPOSPLUGEDSLSS-LHFPQSPPP	715
QY	121	-----TOPVNIPLYASHGAEDF-----SAETEF-----ENHLSPDPSPEQVA	157
Db	716	EWEDLSPLHPFPQPG--EDFOSSLQSPVSCSSSTLSLQSPFESSQSPGPAQ--S	773
QY	158	PVINLEPVE-----ITASHMTVISPDGLLGGMELASESLTTTELDFVFNENDSAVG	207
Db	774	PL--QREPVSSFFSYTLASLQSHESPQSPPE-----	803
QY	208	SIGGABELLIGSLPLS--VDDVESTISFSGP--SSPETSOSSLIESSPELYKVITSSIDAS	263
Db	804	--GPAQSPLOSPVSSPSTSSLSQSSPVSPPSTSSLSKSSPESPLQSPVTSFSSS	861
QY	264	KRFSPXSRSS-----KSKQSVKTSDAKAPKRTKTPAQPEHVIMEHLQKKRKKLO	315
Db	862	TSLSPFSESSSPVDEYTSSTDLLLESDSITDSESILIESEPLFTYTLDEKVDLARPLL-	920
QY	316	NKNAATRYRMKK	327
Db	921	-----LKYYQVKQ	927

RESULT

ID	R1540	standard; Protein; 1863 AA.
AC	R81540;	
DT	02-OCT-1996	(first entry)
DE	BRCA1 mutant from PM20.	
KE	Cancer therapy; breast and ovarian	
KW	antibody production; germline alter	
KS	gene therapy; protein replacement t	
OS	Homo sapiens.	
FT	Key	Location/Qualifiers
FH	misc_difference 1040	
FT		/note= "S1040N"
PN	W09605306-A2.	
PD	22-FEB-1996.	
PF	11-AUG-1995; U10202.	
PF	12-AUG-1994; US-389221.	
PR	02-SEP-1994; US-300266.	
PR	16-SEP-1994; US-308104.	
PR	29-NOV-1994; US-348824.	
PR	24-MAR-1995; US-409305.	
PR	07-JUN-1995; US-483553.	
PR	07-JUN-1995; US-480784.	
PA	(MYRI-) MYRIAD GENETICS INC.	
PA	(CANC-) CANCER INST.	
PA	(RECH-) CENT RECH DU CHUL.	
PI	Durocher F, Emi M, Nakamura Y, S	
DR	WPI: 96-139702/14.	
DR	N-PSDB: T17502.	
PT	New nucleic acid and polypeptide fo	
PT	gene - for diagnosis and therapy of	
PT	and for diagnosing pre-disposition	
CC	Claim 1; 219pp; English.	
CC	R81483-R81497; R81499-R81546 rep	
CC	encoded by the human breast and ova	
CC	(see R81481 for wild type protein).	
CC	immunogens for antibody production.	
CC	these sequences have at least 1 mut	
CC	the wild type cDNA (see T17438 for	

CC alteration in the wild type BRCA1 gene, a predisposition for breast and
 CC ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from
 CC a tissue sample from a subject has a probe, corresponding to a fragment
 CC of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific
 CC probe for a mutation of it), added to it. The conditions allow for
 CC hybridisation of the probe to the mRNA, and any hybridisation which
 CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample is
 CC isolated, and a shift in electrophoretic mobility of single stranded DNA
 CC from the sample on a non-denaturing polyacrylamide gel indicates a
 CC mutation. These methods of detection can also diagnose a lesion
 CC neoplasia associated with the BRCA1 locus. The methods may be used in
 CC gene therapy, protein replacement therapy and protein mimetics, and may
 CC be used to screen for drugs in cancer therapy.
 SQ Sequence 1863 AA;

Query Match 6.3%; Score 121.5; DB 1; Length 1863;
 Best Local Similarity 19.0%; Pred. No. 0.36;
 Matches 83; Conservative 71; Mismatches 137; Indels 145; Gaps 19;
 QY 13 AREWGLEMPVQTDGQFGLKST-----SRHGDESLSLQ 47
 DB 840 SRTSIMESELDAY--LQNTFKYKRSQFAPFSPNGAEECATFSAHG--SLKKQ 895
 QY 48 PQGATLKLEPFEEDVLGAEMWESSDGLSFLDALGDNHERLHPFESNLLEFTSLITPDST 107
 DB 896 SPKVTECEQKEEN-----QGNESNIKPVQT----- 922
 QY 108 VSKDILSLTQFT-----QPVNIPLYASHGAEDFSNETEF---ENHLSPD-----SP 153
 DB 923 -----VNITAGFPVVGOKPVDNAKCSIKGGRFCLSSQFRGNETGLITPNKHGLLQNP 977
 QY 154 EQVAPVNLPEVELTASHMTVPI-----SPDGLLGMELASESLTET-----ELDFV 199
 DB 978 YRTPPLPIKSFVTKCKNLLLENFEHSMSPEREMGNENISTVTSIRNIRENVFK 1037
 QY 200 NFNDSSAVSGTGABELGLSPL-----SVDDVESTIFS-GPS-----SPETSQSS 243
 DB 1038 EANSNNINEVSGTNEVSSINEIGSSDENIAQELGRNRPKLNMLRLGLVQLPEVYKQS 1097
 QY 244 IIESS---PELYK-----VLTSSIDASKRFSYRSRKSQSVKTSIDAKAPKTRTPAQ 295
 DB 1098 LPSGNCKHPKIKQEEVQTVNTD-----FSPYLISDNLEQPMGSSHAS-----QVCS 1147
 QY 296 PVPEHVIMEHLDKDRKKLQKQ---KNAAIRYRMKKKE-----AOGIKGEEQ 339
 DB 1148 ETDDLLDGEIKEDTSAENDIKESSAVFSKVQKELSRSPSPFTHLAQYRGA 1207
 QY 340 ELEELNTKATKVDL 355
 DB 1208 KLESSEENLSSEDEEL 1223

RESULT 6
 ID R13950 standard; Protein; 654 AA.
 AC R13950;
 DT 28-NOV-1991 (first entry)
 DE E2Aalpha protein.
 KW Immunoglobulin; enhancer-binding factor; chromosomal translocation;
 KW leukemia; fusion protein.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT 524..611
 FT /note="region capable of DNA binding and dimer
 FT formation"
 FT
 PN W09113172-A.
 PD 05-SEP-1991.
 PF 22-FEB-1991; U01168.
 PR 23-FEB-1990; US-484063.
 PA (STRD) LELAND STANFORD JR UNIV.
 PA (WHIT-) WHITEHEAD INST BIOMED RE.
 PA Cleary ML, Mellentin JD, Baltimore D, Murre C, Mccaw P;
 DR WPI; 91-281484/38.
 DR N-PSDB; Q13674.
 PT Detection of t(1;19) break-point-associated genes E2A and p1 -
 PT in chromosomal translocation, and prods. useful in diagnosis and
 PT therapy of human neoplasia, esp. acute lymphoblastic leukaemia
 PS Disclosure; Fig 4A; 104pp; English.
 CC An additional Glu390 is present in this protein compared with the
 CC protein encoded by the sequence of Q13670.
 CC A distinctive feature of the N-terminal domain of E2Aalpha is its
 CC high content of proline and glycine.
 CC See also Q13669-75.
 SQ Sequence 654 AA;
 Query Match 6.2%; Score 120.5; DB 1; Length 654;
 Best Local Similarity 20.2%; Pred. No. 0.092;
 Matches 83; Conservative 63; Mismatches 190; Indels 75; Gaps 13;
 QY 4 DLWSEDFQ-----LAREWGLEMPVQTDGQFGLKSTSRHGG---DESLSLQPGQATLK- 54
 DB 226 ELWSPPGQAGFGPMLGGSSPLPLPGSGPVGSSGSSSTFGLHQHERMGYQLHGAEVNG 285
 QY 55 -----LEPFEEDVLGAEMW---ESSLGSLFDALGDNHERLHPFESN 93
 DB 286 GLPSASSFSAPGATYGGVSSHTPPVSGADSLGSRGTTAGSSGDALGKALASIYSPDHS 345
 QY 94 LLEFTSLITPDDSTVSKDILSLTQFTQPVNIPLYASHGAEDFSNETEFENHLSPDSP 153
 DB 346 SNFSS---SPSTPVGSPQGLAGTSQMPRAGAPGALSPSYDGLHGLQSKIEDHL----- 397
 QY 154 EQVAPVNLPEVELTASHMTVPISP-----DGLLGMELASESLTETELDFVNFDSAVG 207
 DB 398 DEATHVLRSHAVG-TAGDMHTLLPGHALASGFTGPMSLGSRH-----AG 441
 QY 208 SIGA--EELLGSLPVDVESTISFGSPSPSTSSSIIESSPELYK-----VISTS 258
 DB 442 LVGGSHPEDGLAGTSLMHNHAAL---PSQPCTLPD---LSRPPDSYSGLRAGATAAAS 495
 QY 259 SIDASKRFSYRSRKSQSVKTSIDAKAPKTRTPAQVPVPEHVIMEHLDKDRKKLQKN 318
 DB 496 EIKREKEDENTISAADHSEEEKELKAPARTSPDEDEDDLLPPPOKAKEREKRVANN 555
 QY 319 AAIIRYRMKKKEAOGIKGEQEELEELNLTUKTKTKVDLQREIKYMKNLMDV 369
 DB 556 ARERLRVRDINEAFKELGRMCQLHNLSEKPTQKLLILHQAVSVILNLEQV 606
 RESULT 7
 ID R81514 standard; Protein; 1863 AA.
 AC R81514;
 DT 02-OCT-1996 (first entry)
 DE BRCA1 mutant from sample set MSK17572.
 KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
 KW antibody production; germline alteration; probe; lesion neoplasia; human;
 KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 1008
 FT /note="M1008I"
 FT
 PN W09605306-A2.
 PD 22-FEB-1996.
 PF 11-AUG-1995; U10202.
 PR 12-AUG-1994; US-289221.
 PR 02-SEP-1994; US-300266.
 PR 16-SEP-1994; US-308104.
 PR 29-NOV-1994; US-348824.
 PR 24-MAR-1995; US-409305.
 PR 07-JUN-1995; US-483553.
 PR 07-JUN-1995; US-480784.
 PA (MYRI-) MYRIAD GENETICS INC.
 PA (CANC-) CANCER INST.
 PA (RECH-) CENT RECH DU CHUL.

PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
 DR WPI; 96-139702/14.
 PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
 PT gene - for diagnosis and therapy of human breast and ovarian cancer
 PT and for diagnosing pre-disposition to these cancers
 PS Claim 1: 218pp; English.
 CC R81483-R81497 and R81499-R81546 represent mutations of the protein
 CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
 CC (see R81481 for wild type protein). These mutations can be used as
 CC immunogens for antibody production. The mutant BRCA1 genes encoding
 CC these sequences have at least 1 mutation or polymorphism in comparison to
 CC the wild type cDNA (see T17438 for wild type). By detecting a germline
 CC alteration in the wild type BRCA1 gene, a predisposition for breast and
 CC ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from
 CC a tissue sample from a subject has a probe, corresponding to a fragment
 CC of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific
 CC probe for a mutation of it), added to it. The conditions allow for
 CC hybridisation of the probe to the mRNA, and any hybridisation which
 CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample is
 CC isolated, and a shift in electrophoretic mobility of single stranded DNA
 CC from the sample on a non-denaturing polyacrylamide gel indicates a
 CC mutation. These methods of detection can also diagnose a lesion
 CC neoplasia associated with the BRCA1 locus. The methods may be used in
 CC gene therapy, protein replacement therapy and protein mimetics, and may
 CC be used to screen for drugs in cancer therapy.
 SQ Sequence 1863 AA;

Query Match 6.2%; Score 120.5; DB 1; Length 1863;
 Best Local Similarity 19.4%; Pred. No. 0.44;
 Matches 85; Conservative 69; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVQTDGQFDLKT-----SRHGGDESLSLQ 47
 Db 840 SRETSIEMESELDAQY--LQNTFKVSKRQSFAPFSPNGNAEECATFSAHSG--SLKKQ 895
 QY 48 PQGATLKEPFEEDVGAENWESSDLGSFLDAGDNHERLHPFESNLLEFTSLTPDDST 107
 Db 896 SPKVTFCEQKEEN-----QCKNESNIKPVQT-----922
 QY 108 VSKDILSSTLQFPT----QVNIPLVASHGAEDFSAETEF---ENHLSPPD-----SP 153
 Db 923 -----VNITAGFPVVGOKDPVDNAKSIKGRFCLSSQFRGNETGLTPNKHGLLQNP 977
 QY 154 EQVAPVINLEPVELTASHMTV-----ISPDGLLGGMELASESLT-----FT 194
 Db 978 YRIPPLFPKSFVKTKCKNLLNEENFEHSISPEREMGNENIPSTVTSIRNNIRENVFK 1037
 QY 195 ELDFVNFND--SAVGSIGGAEELGSLPSVDDVESTISFS-GPS-----SPETSQ 241
 Db 1038 EASSSNIENVGSSSTNEVGSSINEIGS--SDENIQAEIGLRNGRPKLNAMRLGVLPQPEYVK 1095
 QY 242 SSIIESS---PELYK-----VISTSSIDASKRSPYSRKSQSVKTSDAKAPRKTTP 293
 Db 1096 QSLPGSKCKPEAKKQYEEVQVTFD-----FSPYIISLNLEQPMGSSHAS-----QV 1145
 QY 294 AQPVPHEVIMHELDKDRKKLQN---KNAAIRYRMKKGE-----AQIKGE 337
 Db 1146 CSETPDLDLDDGEKEDTSAENDIKESSAVFSKSVQKGLSRSPSPFTTHLAQGYRG 1205
 QY 338 EQELEELNTKLTKVDDL 355
 Db 1206 AKKLESEENLSDEEL 1223

RESULT 8

R81543
 ID R81543 standard; Protein; 1863 AA.

AC R81543;

DT 02-OCT-1996 (first entry)

DE BRCA1 mutant from PM24.

KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;

KW antibody production; germline alteration; probe; lesion neoplasia; human;
 KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
 OS Homo sapiens.

FH Key Location/Qualifiers

FT misc_difference 1008 /note= "M1008I"

FT W09605306-A2.

PD 22-FEB-1996.

PD 11-AUG-1995; U10202.

PF 12-AUG-1994; US-289221.

PR 02-SEP-1994; US-300266.

PR 16-SEP-1994; US-308104.

PR 29-NOV-1994; US-348824.

PR 24-MAR-1995; US-409305.

PR 07-JUN-1995; US-483553.

PR 07-JUN-1995; US-480784.

PA (MYRI-) MYRIAD GENETICS INC.

PA (CANC-) CANCER INST.

PA (RECH-) CENT RECH DU CHUL.

PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;

DR N-PSDB; T17506.

DR WPI; 96-139702/14.

PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1

PT gene - for diagnosis and therapy of human breast and ovarian cancer

PT and for diagnosing pre-disposition to these cancers

PS Claim 1: 218pp; English.

CC R81483-R81497 and R81499-R81546 represent mutations of the protein

CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)

CC (see R81481 for wild type protein). These mutations can be used as

CC immunogens for antibody production. The mutant BRCA1 genes encoding

CC these sequences have at least 1 mutation or polymorphism in comparison to

CC the wild type cDNA (see T17438 for wild type). By detecting a germline

CC alteration in the wild type BRCA1 gene, a predisposition for breast and

CC ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from

CC a tissue sample from a subject has a probe, corresponding to a fragment

CC of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific

CC probe for a mutation of it), added to it. The conditions allow for

CC hybridisation of the probe to the mRNA, and any hybridisation which

CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample is

CC isolated, and a shift in electrophoretic mobility of single stranded DNA

CC from the sample on a non-denaturing polyacrylamide gel indicates a

CC mutation. These methods of detection can also diagnose a lesion

CC neoplasia associated with the BRCA1 locus. The methods may be used in

CC gene therapy, protein replacement therapy and protein mimetics, and may

CC be used to screen for drugs in cancer therapy.

CC Sequence 1863 AA;

Query Match 6.2%; Score 120.5; DB 1; Length 1863;
 Best Local Similarity 19.4%; Pred. No. 0.44;
 Matches 85; Conservative 69; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVQTDGQFDLKT-----SRHGGDESLSLQ 47
 Db 840 SRETSIEMESELDAQY--LQNTFKVSKRQSFAPFSPNGNAEECATFSAHSG--SLKKQ 895
 QY 48 PQGATLKEPFEEDVGAENWESSDLGSFLDAGDNHERLHPFESNLLEFTSLTPDDST 107
 Db 896 SPKVTFCEQKEEN-----QCKNESNIKPVQT-----922
 QY 108 VSKDILSSTLQFPT----QVNIPLVASHGAEDFSAETEF---ENHLSPPD-----SP 153
 Db 923 -----VNITAGFPVVGOKDPVDNAKSIKGRFCLSSQFRGNETGLTPNKHGLLQNP 977
 QY 154 EQVAPVINLEPVELTASHMTV-----ISPDGLLGGMELASESLT-----FT 194
 Db 978 YRIPPLFPKSFVKTKCKNLLNEENFEHSISPEREMGNENIPSTVTSIRNNIRENVFK 1037
 QY 195 ELDFVNFND--SAVGSIGGAEELGSLPSVDDVESTISFS-GPS-----SPETSQ 241
 Db 1038 EASSSNIENVGSSSTNEVGSSINEIGS--SDENIQAEIGLRNGRPKLNAMRLGVLPQPEYVK 1095
 QY 242 SSIIESS---PELYK-----VISTSSIDASKRSPYSRKSQSVKTSDAKAPRKTTP 293

Db 1096 QSLPGSCKPEIKKQBYEEVQVNTD-----FSPYLISNLEQPMGSSHAS-----QV 1145
 QY 294 AQPVPHEVIMHELDKDKRKLQN---KNAAIRYRMKKKE-----AQTGKE 337
 Db 1146 CSETPDLDLGGKEKETSFAENDIKRESSAVFSKSVOKGELSRSPFTTHLAQGYRRG 1205
 QY 338 EQELEELNTKLTKVDDL 355
 Db 1206 AKKLESSEENLSSEDEL 1223

RESULT 9

W48845 ID W48845 standard; Protein; 1503 AA.
 AC W48845;
 DT 12-OCT-1998 (first entry)
 DE Human receptor tyrosine kinase LMR2_h.
 KW Receptor tyrosine kinase; LMR2_h; human; signal transduction;
 KW cancer; neurodegenerative disorder.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..23
 FT /label= Sig_peptide
 FT Domain 43..69
 FT /label= TMD
 FT /note= "transmembrane domain"
 FT Modified_site 409
 FT /label= O-phosphorylated
 FT Modified_site 477
 FT /label= O-phosphorylated
 FT Modified_site 552
 FT /label= O-phosphorylated
 FT Modified_site 1032
 FT /label= O-phosphorylated
 FT Modified_site 1100
 FT /label= O-phosphorylated
 FT Modified_site 1105
 FT /label= O-phosphorylated
 FT Modified_site 1365
 FT /label= O-phosphorylated
 FT Peptide 84..98
 FT /note= "immunogen"
 FT Peptide 684..698
 FT /note= "immunogen"
 FT Peptide 1098..1113
 FT /note= "immunogen"
 FT Peptide 1488..1503
 FT /note= "immunogen"
 FT Peptide 163..166
 FT /note= "protein kinase conserved motif"
 FT Peptide 281..283
 FT /note= "protein kinase conserved motif"
 FT Peptide 329..332
 FT /note= "protein kinase conserved motif"
 PN W09822507-A2.
 PD 28-MAY-1998.
 PE 21-NOV-1997; U22526.
 PR 22-NOV-1996; US-031675.
 PA (SUGEN-) SUGEN INC.
 PI John KE. Plowman CD;
 DR WPI; 98-312419/27.
 DR N-PSDB; V32452.
 PT New isolated receptor tyrosine kinase genes - which are expressed in
 PT neuronal tissues and tumour cells, useful as targets for
 PT neurodegenerative disorders or cancers
 PS Example 1; Fig 7B; 87pp; English.
 CC This is the amino acid sequence of human LMR2-h, deduced from a
 CC LMR2-h cDNA clone (see V32452). Novel rat, human and mouse LMR1,
 CC LMR2 and LMR3 (see W48841-49) define a novel family of receptors
 CC that are structurally related to receptor tyrosine kinases (RTKs).
 CC They all share the distinct motifs that typically characterise this
 CC class of enzymes, but possess extremely short extracellular domains

CC and have C-terminal tails of unprecedented length among RTKs.
 CC Expression of LMR1 and LMR3 is highly restricted to neuronal
 CC tissues with minimal expression in other adult or embryonic organs
 CC or in human tumour cell lines. LMR2 expression is limited to adult
 CC neuronal tissues, but is also very abundantly expressed in other
 CC non-neuronal foetal tissues and in numerous tumour cell lines.
 CC Based on restricted expression of all 3 LMRs to adult neuronal
 CC tissues and the up regulation of LMR2 in a wide variety of tumour
 CC cell lines, these proteins may be critical targets for
 CC neurodegenerative disorders or cancer. Anti-LMR antibodies may be
 CC used for detecting neurodegenerative diseases or cancer
 SQ Sequence 1503 AA;

Query Match 6.2%; Score 119.5; DB 1; Length 1503;

Best Local Similarity 19.5%; Pred. NO. 0.38;

Matches 93; Conservative 73; Mismatches 151; Indels 161; Gaps 24;

QY 11 QLARENGLEMPVVTGQFGD-----LSTSTSYF-----VEVFESSLSDPGKQDDSGQDVP 530
 Db 361 QVIRERDTKLKPKOLEQPYSDRWYEVLOFCWLSPEKRPAAEDVHRLTYLRLOSQR--DS 418
 QY 43 SLSLQPOGATLK-----LEPFEEDVIGAEWMSSDGLGFLDALG-----81
 Db 419 EVDFEQOWNALKPNTNSRDSNNAAFPILDFARDRGLGREMEEVLTVTETSQGLSFEYVW 478
 QY 82 -----DNHERLHPFESNLEFTSLITPDDSTVSKDILSSTLOFP-----120
 Db 479 EAAKHDFERSRGHLDG--LSYTSIFP-----VEVFESSLSDPGKQDDSGQDVP 530
 QY 121 -TQPVNIPLYASHG---AEDFSAETEFENHLS-----PP-----DSPEQVAPVINLEP 164
 Db 531 LRVPGVVPVFDHNLVSGSDYIQLKEKSGSNLELDYPPALLTTDNDNPERTGPELS---587
 QY 165 VELTASHMTVISPDGLLGGMELASESLTFTLDFV--NFNDSAVGSGIGABELGSLPSVD 223
 Db 588 -QLTA-----LRSVELEESS---TDEFFOSSTDPKDSLFG-----620
 QY 224 DVESTISFSGPSP-----ETSSQSSIIESSPELYKVIYSTSIDASKRFSYRSS---K 274
 Db 621 DLHVT---SGPSPFFNFNDVDKSEDLPSHKIFDLMELNGVQAD---FKPATLSSSLDN 675
 QY 275 SKOSVKTSS--DAKAPKRTTPAQVPVPEHVIHEHLDKRKKLQNKNAAIIRYMKKKGEAQ 332
 Db 676 PKESVITGHPEKEKPKRI-FDSEPLCLSDNLHQDNFDFPLNVQELSENFLF-LQEKNLK 733
 QY 333 GIKGEQEELEELNTKLTK-----VDDIQRKIKYMKN-----LMEDVCKANG 374
 Db 734 GSLSSKEHINDLQTELKNAGFTAMLETSCRNLSLDTLOLQFAENKPKGLSLLOENYSTKG 791

RESULT 10

R81523 ID R81523 standard; Protein; 1261 AA.

DT 02-OCT-1996 (first entry)
 DE BRCA1 mutant from sample set TEXAS family BC215-000.
 KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
 KW antibody production; germline alteration; probe; lesion neoplasia; human;
 KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
 OS Homo sapiens.
 PN W09605306-A2.
 PD 22-FEB-1996.
 PF 11-AUG-1995; U10202.
 PR 12-AUG-1994; US-289221.
 PR 02-SEP-1994; US-300286.
 PR 16-SEP-1994; US-308104.
 PR 29-NOV-1994; US-348824.
 PR 24-MAR-1995; US-409305.
 PR 07-JUN-1995; US-483553.
 PR 07-JUN-1995; US-480784.
 PA (MYRI-) MYRIAD GENETICS INC.
 PA (CANC-) CANCER INST.

(RECH-) CENT RECH DU CHUL.
Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
WPI: 96-139702/14.
N-PSDB: T17480.
PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
PS Claim 1; 218pp; English.
CC R81483-R81497 and R81499-R81546 represent mutations of the protein
CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
CC (see R81481 for wild type protein). These mutations can be used as
CC immunogens for antibody production. The mutant BRCA1 genes encoding
CC these sequences have at least 1 mutation or polymorphism in comparison to
CC the wild type cDNA (see T17438 for wild type). By detecting a germline
CC alteration in the wild type BRCA1 gene, a predisposition for breast and
CC ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from
CC a tissue sample from a subject has a probe, corresponding to a fragment
CC of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific
CC probe for a mutation of it), added to it. The conditions allow for
CC hybridisation of the probe to the mRNA, and any hybridisation which
CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample is
CC isolated, and a shift in electrophoretic mobility of single stranded DNA
CC from the sample on a non-denaturing polyacrylamide gel indicates a
CC mutation. These methods of detection can also diagnose a lesion
CC neoplasia associated with the BRCA1 locus. The methods may be used in
CC gene therapy, protein replacement therapy and protein mimetics, and may
CC be used to screen for drugs in cancer therapy.
SQ Sequence 1261 AA;

Query Match	6.1%	Score 118.5;	DB 1;	Length 1261;
Best Local Similarity	19.3%	Pred. No. 0.36;		
Matches 84;	Conservative	70;	Mismatches 135;	Indels 149;
	Gaps	20;		

```

QY      13 ARWGLEMPVOTDQGFDGLKST-----SRHGDESLSLQ 47
   :| | :| | :| | :| | :| |
Db     840 SRETSEMESELDAQY--LONTFKVSQRQSFAPSNPGNAEECATFSAHSG--SLKKQ 895

QY     48 PQGATLKLFPEEDVLGAEMMESSDLGSFLDALGDNHERLHPFESNLLEFTSLTPDDST 107
   :| | :| | :| | :| | :| |
Db    896 SPKVTECEQKEEN-----QKNESNIKPVOT-----922

QY    108 VSKDIILSÜTLOPPT-----QPVNPIYLASHGAEDPSAETF---ENHLSPPD----SP 153
   :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db    923 -----VNITAGFPVVQKDQPDVNAKCSIKGGSRFCJSSQPRGNETGLITPNKHGLQN 977

QY    154 EOAVPINLEPVELTASHTVI-----SPDCLLGGMELASESLT-----FT 194
   :| | :| | :| | :| | :| | :| | :| | :| |
Db    978 YRIPPLFIKSFKVKTKCKNNLEEFCHSPREMGENINIPVTSTIRNNIRENVFK 1037

QY    195 ELDFVNFD--SAVGSIIGAEELLGPLSDVDVESTIFS-GPS-----SPETSQ 241
   :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db   1038 EASSSNIINEVGSSSTEVEGSSINEIGS--SDENIQALGRNGPKLNAMRLGVLOQVEYK 1095

QY    242 SSIIESS---PELYK-----VISTSIDASKRFDPYSRSKSKOSVNTSDAKAPKRTRTP 293
   :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db   1096 QSLFGNCXHKHEIQQEYEVVQTNTD--FPSYLI SDNLEQPMGSSHAS-----QV 1145

QY    294 AQVPVHVTHWEHLDKRRKLQN--KNAAIRYMRMKKG-----AQGIKGE 337
   :| | :| | :| | :| | :| | :| | :| | :| |
Db   1146 CSETPDLDLDDGEINKEDTSFAENDIKESSAVFSKVQGCELS RSPFTHTHLAQGYRRG 1205

QY    338 EQELEENTKLTKYDVDDL 355
   :| | :| | :| | :| |
Db   1206 AKKLESSEENLSDEDEL 1223
```

RESULT 11

R81525
ID R81525 standard; Protein; 1363 AA.

AC R81525;
DT 02-OCT-1996 (first entry)

DE BRCA1 mutant from sample set MONTREAL family 183.

KW KW KW OS PN PD PF PR PR PR PR PA PA PA PI DR DR PT PT PT PS CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC SQ

Query Match	6.18;	Score 118.5;	DB 1;	Length 1363;
Best Local Similarity	19.2%;	Pred. No. 0.4;		
Matches 84;	Conservative	70;	Mismatches 135;	
			Indels 149;	Gaps 20;

```

Qy 13 AREWGLEMPVVTQGQFGDKST-----SRHGDESLSLQ 47
   :| | :| | :| |
Db 840 SRETSIEMEESLDAQY--LQNTFKVSKROSFAPSPNGNAEECATFSAHSG--SLKKQ 895

Qy 48 PQGATLKLEPFEEDVLGAEMWESSDLGSFLDALCONHERLHPFSNLLEFTSLTTPDDOST 107
   :| | :| | :| | :| |
Db 896 SPKVTFCEQKEEN-----QGNESNIKPVQT----- 922

Qy 108 VSKDILSSTLQFFT----QPVNITPLYASHCAEDFAETER---ENHLSPDP-----SP 153
   :| | :| | :| | :| |
Db 923 -----VNITAFPVVGOKDPVDNAKCSIKGGSRFCJSSQPRGNETGLITPNKHGLQNP 977

Qy 154 EQAVPINLEPVELTASHMTVI-----SPDGLLGMLASESLT-----FT 194
   :| | :| | :| | :| | :| |
Db 978 YRIPLPFIKFVKTKCKNLEENFEHSMSPERMGNEINPTVSTISRNIRENVFK 1037

Qy 195 ELDEVNFND--SAVGSIGABEELGSLPSVDVVDTSTIFS-GPS-----SPETSQ 241
   :| | :| | :| | :| | :| | :| |
Db 1038 EASSSNINEVGSSTNVEGSSINEIGS--SDENIQAECLRNRGPKLNAMLRLGVQLQPEYVK 1095

Qy 242 SSIIEST---PELYK-----VISTSIDAKRRFSPIYGRSSKSOSVKTSDAKAPKRTRTP 293
   :| | :| | :| | :| | :| |
Db 1096 OSPLPGNCKNIKEIKOEVEEVOIVNTD---FSPYLISDNLEOPMGSASHAS-----OV 1145

```



```

QY 338 EQELELNTKLTAKVDL 355
Db 1206 AKKLESSEENLSDEEL 1223

RESULT 14
W79896
ID W79896 standard; Protein; 1363 AA.
AC W79896;
DE 09-DEC-1998 (first entry)
DT Tumorigenic BRCA1 protein #582.
KW Tumorigenic allele; ovarian; breast cancer susceptibility gene; BRCA1;
KW identification; specific mutation; allele specific antibody; detection;
KW binding assay; treatment.
OS Homo sapiens.
PN US5821328-A.
PD 13-OCT-1998.
PF 02-APR-1997; 825886.
PR 08-DEC-1993; US-163959.
PR 18-APR-1994; US-232535.
PR 20-OCT-1994; US-326983.
PR 19-APR-1995; US-425061.
PR 02-APR-1997; US-825886.
PA (REGC ) UNIV CALIFORNIA.
PI Friedman L, King M, Lee M, Lynch E, Ostermeyer B,
PI Rowell S, Szabo C;
PI WPI; 98-567730/48.
DR N-PSDB; V60570.
PT Poly(peptide(s) based on mutated BRCA1 allelic sequences - useful
PT for identification of specific allelic mutation(s) of the gene
PT causing breast cancer
PS Claim 1; Columns 105-112; 62pp; English.
CC The present sequence represents a tumorigenic protein of the breast and
CC ovarian cancer susceptibility gene (BRCA1). Exon 11 of the present allele
CC has a 4 bp deletion, which results in a truncated protein. Several
CC other mutagenic alleles have also been identified (see V60560-71). The
CC polypeptides are useful for the identification of specific mutations
CC of BRCA1. The alleles are specific mutations of the C-terminal end of
CC BRCA1, and can be used to compare with the translation product from a
CC patient to identify the mutation that is causing the disease. In this
CC respect, they can also be used to raise allele specific antibodies.
CC They can also be used for detection purposes. The antibodies can be
CC used in binding assays e.g. ELISA. The polypeptides can also be used
CC for the treatment of breast cancer.
SQ Sequence 1363 AA;

Query Match 6.1%; Score 118.5; DB 1; Length 1363;
Best Local Similarity 19.2%; Pred. No. 0.4;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVQDTGQFGLKST-----SRHGDESLSLQ 47
Db 840 SRETSIEMESELDAQY--LQNTFKVSKRQSFAPFNPNGNAEECATFSAHG--SLKKQ 895
QY 48 PQGATLKLEFEEVDLGAEMWESSDLGSLDALGDHNRHLHPFESNLFTSLTPDDST 107
Db 896 SPKVTECEQKEEN-----QGNESNKPQVT-----922
QY 108 VSKDILSLTLOFT-----QPVNIPLYASHGAEDFAETEF---ENHLSPPD-----SP 153
Db 923 -----VNITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRGNETGLITPNKHGLQNP 977
QY 154 EOAPVYNLPVELTASHMTVI-----SPDGLLGGMELASESLT-----ET 194
Db 978 YRIPPLFPKPSFKTKCKMLLENFEHSMSPERMGNIPISTVTSIRNNIRENVFK 1037
QY 195 ELDFVFNFD--SAVSGIGGAELLSGLPSYDDVESTISFS-GFS-----SPETSQ 241
Db 1038 EASSSNINEVGSSTNEVGSSINEIGS--SDENIQAEIGLRNRPKPLNMLRLGLVQPEYK 1095
QY 242 SSIIESS---PELYK-----VISTSSIDASKRFSPIYSRSSKSKQSVKTSDAKAPRTRTP 293

```

```

Db 1096 QSLPGSNCKHPEIKKQYEEVVQTVNTD-----FSPYLIISNLEQPMGSSHAS-----QV 1145
QY 294 AQPVPHEVIMEHLDKDKRKKLQN---KNAATRYRMKKGE-----AQGIKGE 337
Db 1146 CSETPDDLLDDGEIKEDTSFAENDIKESSAVFSKQKGLSRSPSPFTHTHLAQGYRRG 1205
QY 338 EQELELNTKLTAKVDL 355
Db 1206 AKKLESSEENLSDEEL 1223

RESULT 15
R81526
ID R81526 standard; Protein; 1442 AA.
AC R81526;
DT 02-OCT-1996 (first entry)
DE BRCA1 mutant from sample set STRANG family 1900.
KW Cancer therapy; breast and ovarian cancer predisposing gene; Immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
OS Homo sapiens.
PN W09605306-A2.
PD 22-FEB-1996.
PF 11-AUG-1995; U10202.
PR 12-AUG-1994; US-289221.
PR 02-SEP-1994; US-300266.
PR 16-SEP-1994; US-308104.
PR 29-NOV-1994; US-348824.
PR 24-MAR-1995; US-409305.
PR 07-JUN-1995; US-483553.
PR 07-JUN-1995; US-480784.
PA (MYRI-) MYRIAD GENETICS INC.
PA (CANG-) CANG INST.
PA (RECH-) CENT RECH DU CHUL.
PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
PI WPI; 96-139702/14.
DR N-PSDB; T17483.
PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
PS Claim 1; 218pp; English.
CC R81483-R81497 and R81499-R81546 represent mutations of the protein
CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
CC (see R81481 for wild type protein). These mutations can be used as
CC immunogens for antibody production. The mutant BRCA1 genes encoding
CC these sequences have at least 1 mutation or polymorphism in comparison to
CC the wild type cDNA (see T17438 for wild type). By detecting a germline
CC alteration in the wild type BRCA1 gene, a predisposition for breast and
CC ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from
CC a tissue sample from a subject has a probe, corresponding to a fragment
CC of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific
CC probe for a mutation of it), added to it. The conditions allow for
CC hybridisation of the probe to the mRNA, and any hybridisation which
CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample is
CC isolated, and a shift in electrophoretic mobility of single stranded DNA
CC from the sample on a non-denaturing polyacrylamide gel indicates a
CC mutation. These methods of detection can also diagnose a lesion
CC neoplasia associated with the BRCA1 locus. The methods may be used in
CC gene therapy, protein replacement therapy and protein mimetics, and may
CC be used to screen for drugs in cancer therapy.
SQ Sequence 1442 AA;

Query Match 6.1%; Score 118.5; DB 1; Length 1442;
Best Local Similarity 19.2%; Pred. No. 0.44;
Matches 84; Conservative 70; Mismatches 135; Indels 149; Gaps 20;

QY 13 AREWGLEMPVQDTGQFGLKST-----SRHGDESLSLQ 47
Db 840 SRETSIEMESELDAQY--LQNTFKVSKRQSFAPFNPNGNAEECATFSAHG--SLKKQ 895
QY 48 PQGATLKLEFEEVDLGAEMWESSDLGSLDALGDHNRHLHPFESNLFTSLTPDDST 107

```

```
Db 896 SPKVTECEQKEEN-----QKNESENIKPVQT----- 922
QY 108 VSKDILSSTIQFPT-----QPVNIPLYASHCAEDFSAETEF---ENHLSPPD-----SP 153
Db 923 -----VNITAGFPVVGOKDPVDNAKCSIKGSRFCLSSQFRGNETGLITPNKHGLLQNP 977
QY 154 EQVAPVINLEPVELTASHMTVI-----SPDGLGGMELASESLT-----FT 194
Db 978 YRIPPLPIKSFVKTKCKKNLLENFEEHSMSPEREMGNENIPSTVSTISRNNIRENVEK 1037
QY 195 ELDFVNFND--SAVGSIGGAELLGSLVDVDDVESTISFS--GPS-----SPETSQ 241
Db 1038 EASSNINEVGSSTNEVGSSINEIGS--SDENIQAEIGRNRGPKLNAMRLGLVQLPEVYK 1095
QY 242 SSIIESS---PELYK-----VISTSSIDASKRFSFYSRKSQSVKTSDAKAPKTRTP 293
Db 1096 QSLPGSNCKHPEIKKQYEEVQTVNTD----FSPYLISDNLEQPMGSSHAS-----QV 1145
QY 294 AQPVEHVIMEHLDKDKRKLQN---KNAAIRYRMKKKGE-----AQGIKE 337
Db 1146 CSETPDDLDDGEIKEDTSAENDIKESSAVFSVKQKGLSRSPSPFTHLAQGYRRG 1205
QY 338 EQELEELNTKIKTKYDDL 355
Db 1206 AKKLESSEENLSSEDEEL 1223
```

Search completed: August 8, 2000, 02:33:45
Job time: 11669 sec

OM of: US-08-656-811A-1 to: GenEmbl:* out_format : pfs

Date: Aug 8, 2000 6:43 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frames_p2n.model -DEV=xlp
-Q/cgn2.1/USPTO_spool/US08656811/runat_25072000_12317.17103/app_query.fasta_1.443
-DB=GenEmbl -QFMT=fastap -SURFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPEXT=6.000
-FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFM=pfs -NORM=ext -MINLEN=0
-MAXLEN=100000 -USER=US08656811 @CNG1_1_3645 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-08-656-811A-1

Query length: 379

Database: GenEmbl.*

Database sequences: 972840

Database length: 892348106

Search time (sec): 944.170000

score_list:

Sequence	Strd Orig	EScore	Len	Documentation
gb_in1:ACU40851	+ 1923.00	2555.45	1336	! ACU40851 Aplysia californica bZIP
gb_pr1:HUMCREB2A	+ 264.00	344.79	1241	! M86842 Human cAMP response ele
gb_pr1:HUMTAXREBA	+ 264.00	339.91	2015	! D90209 Human mRNA for DNA bind
gb_htg9:AC015801	+ 262.00	232.20	176297	! AC015801 Homo sapiens chrom
gb_htg9:AC021195	+ 262.00	290.44	4.4e-08	209861 ! AC021195 Homo sapiens chro
gb_pr2:HS1104E15	+ 261.00	295.39	2.3e-08	112460 ! AL022312 Human DNA sequen
gb_TO:MUSTRANFAC	+ 259.50	338.93	8.8e-11	1224 ! L13791 Mus musculus (clone C
gb_TO:MUSMATE4A	+ 259.50	337.64	1.0e-10	1391 ! M94087 M. musculus MATE4 (mTR
gb_TO:MWATF4	+ 255.00	331.83	2.2e-10	1365 ! X61507 M. musculus ATF-4 gene
gb_ov:AB013138	+ 246.50	319.82	1.0e-09	1461 ! AB013138 Gallus gallus mRNA fo
gb_pr5:AC002036	+ 231.00	258.72	2.6e-06	80920 ! AC002036 Homo sapiens Chrom
gb_pr1:AB012277	+ 222.00	252.27	3.5e-08	880 ! AB012277 Mus musculus mRNA fo
gb_pr2:HSU03712	+ 220.00	284.63	9.3e-08	1442 ! U03712 Human TAXREB67 pseudog
gb_pr1:AB021663	+ 188.50	242.20	2.1e-05	1506 ! AB021663 Homo sapiens mRNA f
gb_htg9:AC011452	+ 188.50	196.03	0.0080	147307 ! AC011452 Homo sapiens chrom
gb_pr4:AF010388	+ 186.00	245.44	1.4e-05	784 ! AF010388 Homo sapiens activatin
gb_htg5:AC014741	+ 184.00	200.73	0.0044	50935 ! AC014741 Drosophila melanoga
gb_in2:AC005130	+ 184.00	195.65	0.0084	84395 ! AC005130 Drosophila melanoga
gb_in2:AE003670	+ 184.00	192.32	0.0129	117404 ! AE003670 Drosophila melanoga
gb_htg18:AC007208	+ 184.00	189.54	0.0184	154690 ! AC007208 Drosophila melanoga
gb_htg9:AC021154	+ 180.50	183.41	0.0403	178857 ! AC021154 Homo sapiens chrom
gb_htg9:AC036188	+ 171.50	181.88	0.0490	63276 ! AC036188 Homo sapiens chrom
gb_htg4:AC005121	+ 171.00	177.48	0.0862	91685 ! AC005121 Drosophila melanoga
gb_TO:AB012276	+ 162.00	212.64	0.0009	850 ! AB012276 Mus musculus mRNA fo
gb_htg10:AC021163	+ 154.50	148.92	3.36	175994 ! AC021163 Homo sapiens clone
gb_pr3:HUMCREBPAG	+ 145.00	179.98	0.0626	2293 ! L05913 Human cAMP responsive e
gb_pr3:HUMCREBPAB	+ 145.00	179.87	0.0835	2319 ! L05913 Human cAMP responsive e
gb_pr1:HUMCREBPA	+ 145.00	178.57	0.0750	2637 ! L05515 Homo sapiens cAMP resp
gb_pr5:AF197018	+ 143.00	186.39	0.0275	931 ! AF197018 Pan troglodytes HTF1-1
gb_pr3:HUMCREBPAD	+ 142.50	178.03	0.0804	2000 ! L05912 Human cAMP responsive e
gb_TO:RATRAF2	+ 141.00	181.10	0.0542	1209 ! M65148 Rat RAF2 mRNA, complet
gb_pr1:HSCREBP1	+ 140.50	177.32	0.0981	1647 ! X15875 Human mRNA for cAMP res
gb_TO:SF76659	+ 139.50	178.35	0.0771	1302 ! SF76659 cyclic AMP response ele
gb_TO:SF76657	+ 139.50	176.30	0.1003	1596 ! SF76657 cyclic AMP response ele
gb_TO:RNU38938	+ 139.50	174.80	0.1216	1852 ! U38938 Rattus norvegicus cAMP
gb_pr1:HSATFAL	+ 137.50	174.36	0.1287	1485 ! X57197 H.sapiens mRNA for ATF
gb_ov:GGAL7724	+ 137.00	171.69	0.1812	1812 ! Y17724 Gallus gallus mRNA for
gb_ov:XUJ16158	+ 136.50	172.79	0.1573	1520 ! U16158 Xenopus laevis activati
gb_TO:SF76655	+ 135.50	169.49	0.2403	1848 ! SF76655 cyclic AMP response ele
gb_TO:AF062402	+ 134.50	154.25	1.70	7348 ! AF062402 Rattus norvegicus ver
gb_pr12:SPAC25G10	+ 134.50	144.31	6.07	19724 ! 270691 S.pombe chromosome 1 d

```
gb_pl2:S49588 + 134.00 163.31 0.5311 2800 ! S49588 ACR1-ATF/CREB trans
gb_pl1:SCSK01 + 134.00 162.61 0.5806 3000 ! X67875 S.cerevisiae SK01 ge
gb_pl1:SCYNL167C - 134.00 161.32 0.6849 3409 ! Z71443 S.cerevisiae chromos
gb_pl1:SCORFSC14 - 134.00 137.35 14.81 36813 ! X92517 S.cerevisiae PSD1,
```

seq_name: gb_in1:ACU40851

seq_documentation_block:

```
LOCUS ACU40851 1336 bp mRNA 16-MAR-1996
DEFINITION Aplysia californica bZIP transcription factor (ApCREB2) mRNA,
complete cds.
```

ACCESSION U40851

VERSION U40851.1 GI:1123036

KEYWORDS California sea hare.

SOURCE Aplysia californica

ORGANISM Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia;

Anaspidea; Aplysiidae; Aplysia.

REFERENCE 1 (bases 1 to 1336)

AUTHORS Bartsch,D., Ghirardi,M., Skehel,P.A., Karl,K.A., Herder,S.P.,

Chen,M., Bailey,C.H. and Kandel,E.R.

TITLE Aplysia CREB2 represses long-term facilitation: relief of

repression converts transient facilitation into long-term

functional and structural change

JOURNAL Cell 83 (6), 979-992 (1995)

MEDLINE 96107336

REFERENCE 2 (bases 1 to 1336)

AUTHORS Bartsch,D.

TITLE Direct Submission

JOURNAL Submitted (17-NOV-1995) Dusan Bartsch, Howard Hughes Medical

Institute, Columbia University, 722 W. 168th St., New York, NY

10032, USA

FEATURES

Source

1..1336

/organism="Aplysia californica"

/db_xref="taxon:8500"

/tissue_type="central nervous system"

190..1326

/gene="ApCREB2"

190..1326

/gene="ApCREB2"

/codon_start=1

/function="bZIP transcription factor; repressor"

/product="ApCREB2"

/protein_id="AAA92437.1"

/db_xref="GI:1123037"

/translation="MEDLWSEFQLAREWGLEMPVVTQCFQGLKSTRHGCDSEI

SLOPQATLKLEFDEEDVLGAEMWESSDLGSLDALGDNHERLHPFESNLEFTSLIT

PDDSTVSKDLSLTLOPQTQVNIPIYASHGAEDFSAETEFNHLSPDPEQVAPVI

NLEPVELTASHMTVIGPDGLGGMELASELSLTFELDFVFNDSAVSGIGABELLGS

PLSVDDVVESTISFGSPSPSTSSIIIESSEPELYKVISTSSIDASKRFSYSSKSK

QSVKTSDAKAPKTRTPAQPVPBHVMEHLDDKDKKLONKNAIRYMKMKGEAQGI

KGEQEELLENLTKYKVDLQDLQKLMEDVCKAKGILQK"

BASE COUNT 375 a 307 c 328 g 326 t

ORIGIN

alignment_scores:

Quality: 1923.00 Length: 378

Ratio: 5.087 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-08-656-811A-1 x ACU40851 ..

Align seg 1/1 to: ACU40851 from: 1 to: 1336

1 MetGluLeuAspLeuTrpSerGluAspPheGlnLeuAlaArgGluTrpG1 17

|||||

190 ATGGAGCTGGACCTTGGAGCCGAAGATTTCACCTGGCCGAGGATGGG 239

|||||

17 yLeuGluMetProValValGlnThrAspGlyGlnPheGlyAspLeuLysS 34

|||||

|||||

```

240  GCTGGAATGCGAGTCGTCCAGACCGCATGGCCAGTTCGGTGACCTCAAAAT 289
34  eThrSerArgHisGlyAspGluSerLeuSerLeuGlnProGlnGly 50
|||||
290  CAACCAAGTCGTATGTCGGCGAGCAATCTCTAAGTTTGCAGCCCCAGGCG 339
51  AlaThrLeuLysLeuGluProPheGluGluAspValLeuGlyAlaGluTr 67
|||||
340  GCTACACTGAAGTTGGAACCCCTTTGAGGAAGATGCTCTTGGTGCAAGTG 389
67  pMetGluSerSerAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnH 84
|||||
390  GATGAGTCGTCGCGATCTCGGCTCTTTCTGGATGCTTTGGGTGACAACC 439
84  IsGluArgLeuHisProPheGluSerAsnLeuLeuGluPheThrSerLeu 100
|||||
440  ATGAGCGGCTGCATCCGTCGAGTCAAACTTGCCTGAGTTTACATCTCTG 489
101  IleThrProAspSerThrValSerLysAspIleLeuSerSerThrLe 117
|||||
490  ATCACTCTGATGATTCAGCGGTGTCAAAGGACATTCCTCAGCTCAACTCT 539
117  uGluPheProThrGlnProValAsnIleProLeuTyrrAlaSerHisGlyA 134
|||||
540  TCAGTTTCCAACTCAACCACTGAACATCCCTTATATGCAAGTCATGGGG 589
134  laGluAspPheSerAlaGluThrGluPheGluAsnHisLeuSerProPro 150
|||||
590  CCGAAGATTCTCTGCAGAGCTGAGTTTGAGAACCACTGTGCGCTCCA 639
151  AspSerProGluGlnValAlaProValIleAsnLeuGluProValGluLe 167
|||||
640  GATTCTCGGAGCAGTAGCCCTGTCTATAATCTAGAACCAGTTGAAC 689
167  uThrAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyGlyM 184
|||||
690  CACTCGCAGCCATATGACGGTGATCTCACTGATGGCTGTGGTGGGCA 739
184  etGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsn 200
|||||
740  TGGAACTGGCTTCAGAAAGCTTACATTTACCGAACTAGACTTGTGAAC 789
201  PheAsnAspSerAlaValGlySerIleGlyAlaGluGluLeuLeuGly 217
|||||
790  TTCAATGACAGCTGTTGGTTCAATTGGCGGTGCTGAAGAACTTCTTGG 839
217  ySerProLeuSerValAspValGluSerThrIleSerPheSerGlyP 234
|||||
840  CTCGCCACTGTCACTGATGATGTGGAAGTACAATATCATTTTCAGGTC 889
234  roSerSerProGluThrSerGlnSerSerIleIleGluSerSerProGlu 250
|||||
890  CATGTCGCCAGAACACAGCAGCAGCAGCATCATGATCAAGTCCTGAA 939
251  LeuTyrrLysValIleSerThrSerSerIleAspAlaSerLysArgPheSe 267
|||||
940  TTGTACAAAGTTATCTACCTCGCTCCATTCATGATCATCAAGCGTTTCT 989
267  rProTyrrSerArgSerSerLysSerLysGlnSerValIlyThrSerAsp 284
|||||
990  TCCATATCTCTGTTCCCTCCAAGTCCAAGCAATCTGTCAAGACTTCAGAC 1039
284  laLysAlaProArgLysThrArgThrProAlaGlnProValProGluHis 300
|||||
1040  CTAGGCACCTCTGTAACAGGAGGACACCGCGCGACCTGTGCCAGACAT 1089
301  ValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsnLy 317
|||||
1090  GTCATCATGGAACATTTGGCAAAAAAGGACAGAAAGAACTTCAGAACAA 1139
317  sAsnAlaAlaIleArgTyrrArgMetLysLysGlyGluAlaGlnGlyI 334
|||||
1140  GAATGCTGCCATTAGGTATAGGTATCAAGAAAGAGGGGAGGCTCAGGGCA 1189

```

```

334  leLysGlyGluGluGlnGluLeuGluGluLeuAsnThrLysLeuLysThr 350
|||||
1190  TCARAAGGGAGGACAGCAATTAGAGAACTCAACACAAAGCTTAAGACT 1239
351  LysValAspAspLeuGlnArgGluIleLysTyrrMetLysAsnLeuMetG 367
|||||
1240  AAGGTCGATGACTTGCAAAAGAGAAATCAAGTACATGAAAAATTTAATGA 1289
367  uAspValCysLysAlaLysGlyIleGlnLeuLys 378
|||||
1290  AGATGTTTGAAGCGCAAGAGTATTACGCTTAAA 1323
seq_name: gb_pr1:HUMCREB2A

seq_documentation_block:
LOCUS HUMCREB2A 1241 bp mRNA PRI 01-NOV-1994
DEFINITION Human CAMP response element regulatory protein (CREB2) mRNA,
complete cds.
ACCESSION M86842
VERSION M86842.1 GI:181040
KEYWORDS CAMP responsive element regulatory protein.
SOURCE Homo sapiens (tissue library: of J.Leiden) adult cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1241)
AUTHORS Karpinski,B.A., Morle,G.D., Huggenvik,J., Uhler,M.D. and
Leiden,J.M.
TITLE Molecular cloning of human CREB-2: an ATF/CREB transcription factor
that can negatively regulate transcription from the CAMP response
element
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4820-4824 (1992)
MEDLINE 92279218
FEATURES
source Location/Qualifiers
1..1241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Jurkat"
/cell_type="T lymphocyte"
/dev_stage="adult"
/tissue_lib="of J.Leiden"
/map="2q32"
108..1163
/gene="CREB2"
108..1163
/gene="CREB2"
/codon_start=1
/db_xref="GDB:G00-128-011"
/product="CAMP response element regulatory protein"
/protein_id="AAA52071.1"
/db_xref="GI:181041"
/translation="MTFMSFLSEVLVDLMSPPDFPSPGLGAEESLGLDDYLEVAKHF
KPHGFSDDKAKAGSSEWLAVDLVSPSNNSKEDAFSGTDMLEKMDLKEFDLIDLGI
DDELTMPDLDLTTLDLDTCDLFLAQVETNKQPQVNPVNIHLPESLTKPDQVAPFTFL
QPLPLSPGVLSSTPDHFSFSLGSEVDITEGDKRPDYAYVAMIPQCIKEEDTPSDND
SGICMSPEYLGSPOHSPSTGSPNRSPLSPGVLCGSARPKYPDPGKMWAAKVGKE
KLDKLLKMEQNKTAATRYRQKRAEQEALTGECKELEKKNEALKERADSLAKEIQL
KDLIEVRKARKKRKRP"
polyA_site 1241
/gene="CREB2"
/note="G00-128-011"
BASE COUNT 328 a 310 c 336 g 267 t
ORIGIN Chromosome 2q32.

alignment_scores:
Quality: 264.00 Length: 408
Ratio: 1.294 Gaps: 22
Percent Similarity: 50.000 Percent Identity: 28.676

alignment_block:
US-08-656-811A-1 x HUMCREB2A

```

19 GluMetProValValGlnThrAspGlyGlnPheGlyAspLeuLysSer... 34
114 GAAATCAGCTTCTTCAGCAGCAGGAGCTGTGGTGGGGACACTTGATGTCCCC 163
35ThrSerArgHisGlyGlyAspGluSerLeuSerLeu... 46
164 CTTGCACCCGTCGGGTTGGGGGCTGAAGAAAGCCTAGGTCTCTTAGATG 213
47GlnProGlnGlyAlaThr... 52
214 ATTACCTGGAGGTGCCAAGCACATTCAACACTCATGGTCTCCACGGAC 263
53LeuLysLeuGluProPhe... 58
264 AAGGCTAAGCGCGGCTCCTCGAATGCTGCCTGTGGATGGTGGTGCAG 313
59GluGluAspValLeuGlyAlaGluTrpMet... 68
314 TCCTCCCAACACAGCAGGAGGATGCTTCTCGGGACAGATTGGATGT 363
69 ..GluSerSerAspLeuGlySerPhe...LeuAspAlaLeuGlyAspAsn 83
364 TCGAGAAAATGATTTGAAGCAGGTTCGACTTGGATGCC... 401
84 HisGluArgLeuHisProPheGluSerAsnLeuGluPheThrSerLe 100
402CTGTGGGTATAGATGACCT 421
100 uIleThr...ProAspSerThrValSerLysAsp... 111
422 GGAACCATGCGCAGATGACCTTCGACCATGTTGGATGACACTGTGTATC 471
112IleLeuSerSerThrLeuGlnPheProThrGlnProVal 124
472 TCTTTTGCCCTAGTCCAGGAGCTAATAAGCAGCCGCCACAGCGGC 521
125 AsnIleProLeuTyAlaSerHisGlyAlaGluAspPheSerAlaGluTh 141
522 AAC...CCAATT... 530
141 rGluPheGluAsnHisLeuSerProAspSer...ProGluG 155
531GGCATCTC...CCAGAAAGTTTAACAAAACCCGACC 564
155 InValAlaProValIleAsnLeuGluProValGluLeuThrAlaSerHis 171
565 AGGTTGCCCTTACCTTCTTACAACTCTTCCCTTTCCCGCAGGGTC 614
172 MetThrValIleSerProAspGlyLeuLeuGlyGlyMetGluLeuAla 188
615 CTGTCC...TCCACTCCAGATCATTCCTTT...AGTTTAGAGTGGCGAG 658
188 rGluSerLeuThrPheThrGluLeuAspPheValAsnPheAsnAspSera 205
659 TGA...GTGATATCATCAGGAGATAGAGACGACGACTACACTGCTT 705
205 lValIcylSerIleGlyAlaGluLeuGlySerProLeuSer 221
706 ACCTTGCCATGATC...CCTCAGTC 728
222 ValAspAspValGluSerThrIleSerPheSerGlyProSerSerProG 238
729 ATAAAGGAGGAAGAC...ACCCCTTC 751
238 uThrSerGlnSerSerIleGluSerSerProGluLeuTyLysVal 255
752 AGATAATGATAGTGCATCTCT...ATGACCCAGACTCTATCTG... 794
255 leSerThrSerSerIleAspAlaSerLysArgPhePheProTySerArg 271

```

795 .....GGGTCTCTCAGCAGACAGCCCTCTTACCAGG 824

272 SerSerLysSerLysGlnSerValLysThrSerAspAlaLysAlaProAr 288
   ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
825 GGCTCTCCAAATAGG...AGCTCCCATCTCCAGGTGTCTCTGTGGTCTC 871

288 gLysThrArgThrProAlaGlnProValProGluHisValIleMet.... 303
   :::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
872 TGCCGCTGCCAACCTTACGATCCTCCTGGAGAGAGATGGTAGCAGCAA 921

304 .....GluHisLeuAspLysLysAspArgLysLysLeuGlnAsn 316
   :::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
922 AAGTAAGGCTGAGAACTGGATAGAGCTGAAAAAATGAGGAGCAAAAC 971

317 LysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGluAlaGlnGl 333
   :::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
972 AAGACAGCAGCCACTAGGTACCGCCAGAGAGAGAGCGGCGGACGAGGC 1021

333 yIleLysGlyGluGlnGlnLeuGluGluLeuAsnThrLysLeuLysT 350
   :::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
1022 TCTTACTGCTAGTGCAAGAGCTGGAAAGAGAACGAGGCTCTAAAAG 1071

350 hrLysValAspAspLeuGlnArgGluIleLysTyrMetLysAsnLeuMet 366
   :::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
1072 AGAGCGCGGATTCCTCGCCAAGGAGATCCAGTACCTGAAAGATTGATA 1121

367 GluAspValCysLysAlaLysGly 374
   :::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
1122 GAAGAGGTCGCAAGCAAGCAGGAGGG 1145

seq_name: gb_pr1:HUMTAXREBA

seq_documentation_block:
LOCUS HUMTAXREBA 2015 bp mRNA PFI 07-FEB-1999
DEFINITION Human mRNA for DNA binding protein TAXREB67.
ACCESSION D90209
VERSION D20209.1 GI:220087
KEYWORDS CREB; DNA binding protein; TAX-responsive enhancer.
SOURCE Human fibroblast cell line GM637, cDNA to mRNA, clone pcd67.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 2015)
Tsujimoto,A., Niyunyu,H., Morita,T., Sato,T. and Shimotohno,K.
Isolation of cDNAs for DNA-binding proteins which specifically bind
to a tax-responsive enhancer element in the long terminal repeat of
human T-cell leukemia virus type I
J. Virol. 65 (3), 1420-1426 (1991)
91140735
Submitted (27-AUG-1990) to dbj by:
Kunitada Shimotohno
National Cancer Center Research Institute
5-1-1 Tsukiji, Chuo-ku
Tokyo 104
Japan
Phone: 03-542-2511
Fax: 03-545-3567
Location/Qualifiers
1. .2015
/organism="Homo sapiens"
/db_xref="taxon:9606"
882. .1937
/codon_start=1
/product="TAXREB67 protein"
/protein_id="BAAL4234.1"
/db_xref="GI:220088"
/translators="MTEPSFSLSEVLGMDLKEFDLDALEL
KPHGFSSDKAKAGSEWLVGGLVSPNSNKKEDAFSGTDMLEKMDLKEFDLDALEL
DDLETPDLLTLTDLCDLFAFLVQETNKKPQTVNPGHLESLTKPDQVAPFTFL
QPLPSGVSLSSTSPSHSFLSELGSEVDITGDRKPDYTAIVAMIPQICEEDTPSDN
SCMSPGVSLSQPSHSFSPNSPGLSGSARPKPYDPGPKRMVAAKVKGE
KDLQKLIKMPQNKTAATRYQKKRAQEALTGCEKLEKNEALKERADSLAKEIQIYL
KDLIEYVRKARGKKRPV"

FEATURES
source
CDS

```

```

795 .....GGGTCTCCTCAGCACAGCCCTCTTACCAGG 824

272 SerSerLysSerLysGlnSerValLysThrSerAspAlaLysAlaProAr 288
    ::::| | | | | ::::| | | | | ::::| | | | |
825 GGCTCTCCAATAGG...AGCTCCCATCTCCAGGTGTTCTCTGTGGTC 871

288 gLySthrArgThrProAlaGlnProValProGluHisValIleMet.... 303
    ||| | | | | | | | | | | | | | | | | | |
872 TGCCCGTGCCCAACCTTAGCATCTCCTCGGAGGAAGATGGTAGCGAA 921

304 .....GluHisLeuAspLysLysAspArgLysLysLeuGlnAsn 316
    ::::| | | | | | | | | | | | | | | | | | |
922 AAGTAAGGCTGAGAACAATGGATAGAACGCTCAAATAAATGAGCAAAA 971

317 LysAsnAlaAlaIleArgTyrArgMetLysLysGlyGluAlaGlnGl 333
    ::::| | | | | | | | | | | | | | | | | | |
972 AAGACAGCAGCCACTAGGTACGCCAGAGAGAGAGCGGCAGCAGAGGC 1021

333 ylleLysGlyGluGlnGluLeuGluGluLeuAsnThrLysLeuLysT 350
    ::::| | | | | ::::| | | | | ::::| | | | |
1022 TCTTACTGCTGAGTGCAAGAGCTGGAAAGAAAGAACGAGGCTCTAANA 1071

350 hrLysValAspAspLeuGlnArgGluIleLysTyrMetLysAsnLeuMet 366
    ::::| | | | | ::::| | | | | ::::| | | | |
1072 AGAGCGCGGATTCCCTGGCCAGGAGATCCAGTACCTCAAAGATTGATA 1121

367 GluAspValCysLysAlaLysGly 374
    ::::| | | | | ::::| | | | | ::::| | | | |
1122 GAAGAGGTCGCAAGCAAGCAGGAGGG 1145

seq_name: gb_pr1:HUMTXREBA

seq_documentation_block:
LOCUS HUMTXREBA 2015 bp mRNA PFI 07-FEB-1999
DEFINITION Human mRNA for DNA binding protein TAXREB67.
ACCESSION D90209
VERSION D20209.1 GI:220087
KEYWORDS CREB; DNA binding protein; TAX-responsive enhancer.
SOURCE Human fibroblast cell line GM637, cDNA to mRNA, clone pcd67.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
Tsujimoto,A., Niyunyu,H., Morita,T., Sato,T. and Shimotohno,K.
Isolation of cDNAs for DNA-binding proteins which specifically bind
to a tax-responsive enhancer element in the long terminal repeat of
human T-cell leukemia virus type I
J. Virol. 65 (3), 1420-1426 (1991)
91140735
Submitted (27-AUG-1990) to dbj by:
Kunitada Shimotohno
National Cancer Center Research Institute
5-1-1 Tsukiji, Chuo-ku
Tokyo 104
Japan
Phone: 03-542-2511
Fax: 03-545-3567
Location/Qualifiers
1. .2015
/organism="Homo sapiens"
/db_xref="taxon:9606"
882..1937
/codon_start=1
/product="TAXREB67 protein"
/protein_id="BAAL4234.1"
/db_xref="GI:220088"
/translators="MTEPSFSLSEVLVDLMKEDAFSGTDMLKEMDLKEDLDALLGI
KPHGFSSDKAKAGSEWLVADGLVSPNSNKSDFATSGTDMLKEMDLKEDLDALLGI
DDLETPDLLLLTLTDCLDLFAELVQETNKQPQTVPNGICHLPESLTKPDQVAPFTFL
QPLPSPGVLSUSTSPHSFSLSELGEVDITGDRKKPDYTAIVAMIPQCIEEDTPSDN
SCMGSPSVLSQPSISPSRPNLSPGSLGCSARPKPYDPGPRKMAVKVAGE
KLQDKLIKMPQNKAATRYRQKRRAQEALTGECCKELEKKNALKERADSIAEKIQYL
KDILIEVRKARGKKRPV"

FEATURES
source
CDS

```


TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6563606.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2128
Center clone name: 854_A_13

*** NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 48: contig of 48 bp in length
* 49 148: gap of 100 bp
* 149 1183: contig of 1035 bp in length
* 1184 1283: gap of 100 bp
* 1284 3141: contig of 1858 bp in length
* 3142 3241: gap of 100 bp
* 3242 4983: contig of 1742 bp in length
* 4984 5083: gap of 100 bp
* 5084 7035: contig of 1952 bp in length
* 7036 7135: gap of 100 bp
* 7136 10589: contig of 3454 bp in length
* 10590 10689: gap of 100 bp
* 10690 14230: contig of 3541 bp in length
* 14231 14330: gap of 100 bp
* 14331 17798: contig of 3468 bp in length
* 17799 17898: gap of 100 bp
* 17899 20725: contig of 2827 bp in length
* 20726 20825: gap of 100 bp
* 20826 24122: contig of 3297 bp in length
* 24123 24222: gap of 100 bp
* 24223 27097: contig of 2875 bp in length
* 27098 27197: gap of 100 bp
* 27198 32876: contig of 5679 bp in length
* 32877 32976: gap of 100 bp
* 32977 39229: contig of 6253 bp in length
* 39230 39329: gap of 100 bp
* 39330 44846: contig of 5517 bp in length
* 44847 44946: gap of 100 bp
* 44947 51703: contig of 6757 bp in length
* 51704 51803: gap of 100 bp
* 51804 58441: contig of 6638 bp in length
* 58442 58541: gap of 100 bp
* 58542 65154: contig of 6613 bp in length
* 65155 65254: gap of 100 bp
* 65255 73330: contig of 8076 bp in length
* 73331 73430: gap of 100 bp
* 73431 82270: contig of 8840 bp in length
* 82271 82370: gap of 100 bp
* 82371 92695: contig of 10325 bp in length
* 92696 92795: gap of 100 bp
* 92796 101707: contig of 8912 bp in length
* 101708 101807: gap of 100 bp
* 101808 110885: contig of 9078 bp in length
* 110886 110985: gap of 100 bp
* 110986 123688: contig of 12703 bp in length
* 123689 123788: gap of 100 bp
* 123789 140525: contig of 16737 bp in length
* 140526 140625: gap of 100 bp
* 140626 159784: contig of 19159 bp in length

* 159785 159884: gap of 100 bp
* 159885 176297: contig of 16413 bp in length.
FEATURES
source
1. 176297
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-854A13"
/clone_11b="RPC1-11 Human Male BAC"
1. 48
/note="assembly_fragment
clone_end:17
vector_side:right"
149. 1183
/note="assembly_fragment"
1284. 3141
/note="assembly_fragment"
3242. 4983
/note="assembly_fragment"
5084. 7035
/note="assembly_fragment"
7136. 10589
/note="assembly_fragment"
10690. 14230
/note="assembly_fragment"
14331. 17798
/note="assembly_fragment"
17899. 20725
/note="assembly_fragment"
20826. 24122
/note="assembly_fragment"
24223. 27097
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
27198. 32876
/note="assembly_fragment"
32977. 39229
/note="assembly_fragment"
39330. 44846
/note="assembly_fragment"
44947. 51703
/note="assembly_fragment"
51804. 58441
/note="assembly_fragment"
58542. 65154
/note="assembly_fragment"
65255. 73330
/note="assembly_fragment"
73431. 82270
/note="assembly_fragment"
82371. 92695
/note="assembly_fragment"
92796. 101707
/note="assembly_fragment"
101808. 110885
/note="assembly_fragment"
110986. 123688
/note="assembly_fragment"
123789. 140525
/note="assembly_fragment"
140626. 159784
/note="assembly_fragment"
159885. 176297
/note="assembly_fragment"
BASE COUNT 44519 a 42534 c 43191 g 43536 t 2517 others
ORIGIN
alignment_scores:
Quality: 262.00 Length: 401
Ratio: 1.278 Gaps: 20

Insert size: 225000; agarose-fp
 Insert size: 206861; sum-of-ctnigs
 Quality coverage: 3.50 in Q20 bases; agarose-fp
 Quality coverage: 3.75 in Q20 bases; sum-of-ctnigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 31 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1 2315: contig of 2315 bp in length
* 2316: gap of unknown length
* 2416: contig of 2157 bp in length
* 4572: gap of unknown length
* 4673: contig of 2337 bp in length
* 7009: gap of unknown length
* 7010: gap of unknown length
* 7110: contig of 2656 bp in length
* 9765: gap of unknown length
* 9865: contig of 2608 bp in length
* 12474: contig of 4093 bp in length
* 12574: gap of unknown length
* 16666: gap of unknown length
* 16767: contig of 3247 bp in length
* 20013: gap of unknown length
* 20113: contig of 3895 bp in length
* 24108: gap of unknown length
* 24109: contig of 3941 bp in length
* 28049: gap of unknown length
* 28149: gap of unknown length
* 31940: contig of 3791 bp in length
* 32040: gap of unknown length
* 35923: contig of 3883 bp in length
* 36023: gap of unknown length
* 40326: contig of 4303 bp in length
* 40426: gap of unknown length
* 40427: contig of 4844 bp in length
* 45270: gap of unknown length
* 45370: contig of 3445 bp in length
* 48815: gap of unknown length
* 48916: contig of 4989 bp in length
* 53904: gap of unknown length
* 54004: gap of unknown length
* 58337: contig of 4333 bp in length
* 58437: gap of unknown length
* 63057: contig of 4620 bp in length
* 63157: gap of unknown length
* 67683: contig of 4526 bp in length
* 67783: gap of unknown length
* 74319: contig of 6536 bp in length
* 74419: gap of unknown length
* 83151: contig of 8732 bp in length
* 83251: gap of unknown length
* 89129: contig of 5878 bp in length
* 89229: gap of unknown length
* 98220: contig of 8991 bp in length
* 98320: gap of unknown length
* 105559: contig of 7239 bp in length
* 105659: gap of unknown length
* 113081: contig of 7422 bp in length
* 113181: gap of unknown length
* 120975: contig of 7794 bp in length
* 121075: gap of unknown length
* 132903: contig of 11828 bp in length
* 132904: gap of unknown length
* 144491: contig of 11488 bp in length
* 144591: gap of unknown length
* 153726: contig of 9135 bp in length
* 153826: gap of unknown length
* 153827: contig of 17504 bp in length
* 171331: gap of unknown length
* 171431: contig of 20489 bp in length
* 191919: gap of unknown length
* 192019: gap of unknown length

```

```

* 192020 209861: contig of 17842 bp in length.
FEATURES
  source
    1..209861
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="17"
    /clone="RP11-794C22"
BASE COUNT 52775 a 51286 c 51341 g 51413 t 3044 others
ORIGIN
alignment_scores:
  Quality: 262.00      Length: 401
  Ratio: 1.278        Gaps: 20
  Percent Similarity: 51.122  Percent Identity: 28.180
alignment_block:
  US-08-656-811A-1 x AC021196 ..
Align seg 1/1 to: AC021196 from: 1 to: 209861
19 GluMetProValValGlnThrAspGlyGlnPheGlyAspLeuLysSer... 34
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
196107 GAATGAGCTTCCTGAGCAACGAGGTGTGGTGGGGGACTTGATGTCCCC 196156
35 .....ThrSerArgHisGlyGlyAspGluSerLeuSerLeu..... 46
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
196157 CTTCCAGCAGTCGGGTGGGGGCTGAAGAAAGCATAGTCTCTTTAGATG 196206
47 .....GlnProGlnGlyAlaThrLeuLys 54
196207 ACTAGTGGAGTGCCCAAGCACCTTCAACACCTCATGGG..... 196244
55 LeuGluProPheGluGluAspValLeuGlyAla.....GluTrpMe 68
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
196245 .....TTCCTCAGCAGCAAGGCTAAGCGGGCTCTCCGATGGCT 196285
68 tGluSerSerAspLeu.....GlySerPheLeuAspAlaL 80
: : : : : : : : : : : : : : : : : : : : : : : : : :
196286 GACTGTGGATGGTGGTCAGTCCTCCAAACAACAGCAAGGAGGATGCT 196335
80 euGlyAspAsnHisGluArgLeuHisProPheGluSerAsnLeuGlu 96
: : : : : : : : : : : : : : : : : : : : : : : : : :
196336 TCTCTGGGACACATGGATGTGGAGAAATGGATTTGAAGGAGTTCGAC 196385
97 PheThrSerLeuLeuThrProAspAsp...SerThrValSerLysAspL 112
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
196386 TTTGGTGCCTGTGGGTATAGATGACCTGGAAACCATGCCAGATGACCT 196435
112 eLeuSerSer.....GGCCATCTC... 115
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
196436 TTTGACCACGTTGGATGACACTTGTAACTCTTTGCCCCCTTAGTCCAGG 196485
116 ..ThrLeuGlnPheProThrGlnProValAsnIleProLeuTyrlaser 131
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
196486 AGACTAATAAGGTGCCCCCGCAGATGGTGAAC...CCAATT..... 196523
132 HisGlyAlaGluAspPheSerAlaGluThrGluPheGluAsnHisLeu 148
196524 .....GGCCATCTC... 196532
148 rProProAspSer.....ProGluGlnValAlaProValIleAsnL 162
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
196533 ....CCAGAAAGTTTAAACAAACCGCAGGTTGCCCCCTTCACCTTCT 196578
162 euGluProValGluLeuThrAlaSerHisMetThrValIleSerProAsp 178
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
196579 TGCACACTTCTTCCCCCTTCCCGCAGG...GTCCAGTCTCCACTCCAGAT 196625
179 GlyLeuLeuGlyGlyMetGluLeuAlaSerGluSerLeuThrPheThrCl 195
: : : : : : : : : : : : : : : : : : : : : : : : : :
196626 CATTCCTTT...AGTTTAGAGCTGGGCAAGTAA...GTGGATATCACTGA 196669

```

```

195 uLeuAspPheValAsnPheAsnAspSerAlaValGlySerIleGlyGlyA 212
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
196670 AGAAGATAGGAAGCGGACTCCACTGCTTACGTTGCCATGATC..... 196712
212 laGluGluLeuLeuGlySerProLeuSerValAspValGlySerThr 228
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
196713 .....CCTCAGTGCATAAAGGAGGAAGAC..... 196736
229 IleSerPheSerGlyProSerProGluThrSerGlnSerSerIleI 245
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
196737 .....ACCCCTCAGATAATGATGTCGATCTG 196765
245 eGluSerSerProGluLeuTyrlsValIleSerThrSerSerIleAspA 262
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
196766 T...ATGAGCCACAGTCTATCTG.....G 196788
262 laSerLysArgPheSerProTyrlsSerArgSerSerLysGlnSer 278
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
196789 GGTCTCTCCAGCATAGCCCTCTACCAAGGGCTCTCCAAATAGG...AGC 196835
279 ValLysThrSerAspAlaLysAlaProArgLysThrArgThrProAlaG 295
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
196836 CTCCACTCTCCAGTGTCTCTGTGGTCTGCCACCACCAACCTTAGCA 196885
295 nProValProGluHisValIleMet.....GluHisLeuA 307
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
196886 TCCTCTGGAGAGAAGATGGTAGCAGCAAAAGTAAGGGTCGAGAACTGG 196935
307 spLysLysAspArgLysLysLeuGlnAsnLysAsnAlaIleArgTyr 323
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
196936 ATAAGAAGCTGAAAAAATGGAGCAAAACAGACAGCAGCCACTAGGTAC 196985
324 ArgMetLysLysLysGlyGluAlaGlnGlyIleLysGlyGluGluGln 340
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
196986 CCCAGAGAGAGAGGGGAGCAGCAGAGGCTCTACTGCGAGTGCARA 197035
340 uLeuGluGluLeuAsnThrLysLeuLysThrLysValAspLeuGlnA 357
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
197036 GCTGAAAAGAGAAGAGAGGCTCTAAAAGAGAGGGCAGATCCCTGGCCA 197085
357 rGluIleLysTyrMetLysAsnLeuMetGluAspValCysLysAlaLys 373
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
197086 AGGAGATCCAGTACTGAAAGATTGTAGAGAGGTCCTCCGAGGCAAGG 197135
374 Gly 374
   |||
197136 GGG 197138

```

seq_name: gb_pr2:HS1104E15

seq_documentation_block:
LOCUS HS1104E15 112460 bp DNA PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone RP5-1104E15 on chromosome 22q12.3-13.1. Contains the MGAT3 gene for mannosyl (beta-1,4)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, the gene for a predicted protein, the AT4 gene for activating transcription factor 4 (tax-responsive enhancer element B67) and the 5' end of the CACNA1I gene for voltage-dependent calcium channel, alpha 1I subunit. Contains ESTs, STSs, GSSs and five putative CpG islands, complete sequence.

ACCESSION AL022312
VERSION AL022312.7 GI:4914501
KEYWORDS HTG; AT4; CACNA1I; calcium channel; CpG island; MGAT3.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 112460)
AUTHORS Barlow,K.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

COMMENT

On May 28, 1999 this sequence version replaced gi:490589. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C_elegans/wormpep RP5-1104E15 is from the library RPCI-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP5-1104E15. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone RP5-1104E15 is at 1 in this sequence. The true left end of clone CTA-206C7 is at 112356 in this sequence. The true right end of clone RP3-407F17 is at 83950 in this sequence. The end of this sequence overlaps with sequence AL008716.

FEATURES

Source

1. .112460
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="22"
/map="q12.3-13.1"
/clone="RP5-1104E15"
/clone_lib="RPCI-5"
1. .62
/note="Alu repeat: matches 1. .62 of consensus"
repeat_region
82. .149
/note="MIR repeat: matches 75. .140 of consensus"
repeat_region
150. .440
/note="AluJo repeat: matches 17. .303 of consensus"
repeat_region
441. .556
/note="MIR repeat: matches 140. .262 of consensus"
repeat_region
1072. .1198
/note="MIR repeat: matches 13. .143 of consensus"
repeat_region
1205. .1511
/note="AluSx repeat: matches 1. .308 of consensus"
repeat_region
1751. .1841
/note="MER63 repeat: matches 26. .752 of consensus"
repeat_region
1891. .1931
/note="L2 repeat: matches 2664. .2704 of consensus"
2297. .2644
/note="match: GSS: Em:AQ218889"
repeat_region
2514. .2569
/note="MIR repeat: matches 113. .170 of consensus"
repeat_region
2592. .2887
/note="AluSc repeat: matches 1. .298 of consensus"
repeat_region
3001. .3085
/note="L2 repeat: matches 2619. .2710 of consensus"
repeat_region
3088. .3127
/note="5 copies 8 mer atgaatga 92 conserved"
repeat_region
3839. .4037
/note="MIR repeat: matches 31. .248 of consensus"
repeat_region
4197. .4326
/note="MER5A repeat: matches 20. .149 of consensus"
repeat_region
4343. .4626

```

repeat_region /note="AluJo repeat: matches 25. .305 of consensus"
5079. .5125
/note="L2 repeat: matches 2686. .2735 of consensus"
5126. .5429
/note="AluSx repeat: matches 1. .303 of consensus"
5439. .5743
/note="AluY repeat: matches 1. .300 of consensus"
6365. .6649
/note="AluSx repeat: matches 27. .311 of consensus"
7636. .7708
/note="L2 repeat: matches 2638. .2710 of consensus"
8325. .8386
/note="31 copies 2 mer tt 72 conserved"
8328. .8387
/note="L2 copies 5 mer tcttt 86 conserved"
8412. .8493
/note="Alu repeat: matches 225. .308 of consensus"
9387. .9483
/note="L2 repeat: matches 2611. .2710 of consensus"
10720. .11321
/note="L2 repeat: matches 2065. .2709 of consensus"
11479. .11599
/note="MIR repeat: matches 31. .145 of consensus"
12182. .12267
/note="43 copies 2 mer cc 64 conserved"
13334. .13419
/note="MIR repeat: matches 44. .144 of consensus"
13871. .14283
/note="MSTD repeat: matches 1. .391 of consensus"
complement(15277. .15778)
/note="match: GSS: Em:AQ317473; match: STS: Em:G56767"
13818. .15839
/note="11 copies 2 mer tt 100 conserved"
complement(15839. .16591)
/note="match: GSS: Em:AQ740632"
complement(15846. .16591)
/note="match: GSS: Em:AQ740620"
16460. .16520
/note="MER3 repeat: matches 146. .209 of consensus"
complement(16461. .17266)
/note="match: GSS: Em:AQ739191"
16521. .16647
/note="FLAN_A repeat: matches 1. .133 of consensus"
17352. .17478
/note="MIR repeat: matches 32. .148 of consensus"
18248. .18455
/note="MIR repeat: matches 51. .262 of consensus"
18484. .18655
/note="MIR repeat: matches 84. .256 of consensus"
19753. .20050
/note="AluJo repeat: matches 1. .293 of consensus"
20336. .20350
/note="MIR repeat: matches 246. .259 of consensus"
20351. .20410
/note="L2 repeat: matches 2666. .2729 of consensus"
20411. .20476
/note="MIR repeat: matches 40. .246 of consensus"
20554. .20642
/note="L2 repeat: matches 2531. .2620 of consensus"
20687. .20747
/note="MIR repeat: matches 93. .153 of consensus"
21468. .22247
/note="L2 repeat: matches 1833. .2750 of consensus"
complement(22233. .22544)
/note="match: GSS: Em:AQ594604"
22466. .22617
/note="MIR repeat: matches 82. .262 of consensus"
22640. .22977
/note="AluJb repeat: matches 1. .300 of consensus"
23096. .23274
/note="MIR repeat: matches 10. .192 of consensus"
23902. .24207
/note="AluSp repeat: matches 1. .312 of consensus"

```

```

repeat_region 24743. .24836
/note="MIR repeat: matches 35. .142 of consensus"
24996. .25049
/note="MIR repeat: matches 92. .148 of consensus"
25050. .25361
/note="AluSx repeat: matches 1. .312 of consensus"
25410. .25700
/note="AluSx repeat: matches 1. .291 of consensus"
26006. .26197
/note="L1MC3 repeat: matches 7546. .7734 of consensus"
26198. .26512
/note="AluSx repeat: matches 1. .312 of consensus"
26520. .32881
/gene="MGAT3"
/note="match: CDNAS: Em:E11285; match: ESTs: Em:A1452685
Em:A1524129 Em:A1198445 Em:A1439025 Em:A1075278
Em:AA743691 Em:AA421504 Em:A1090769 Em:A1308859
Em:AA829826 Em:A1571173 Em:AA501937 Em:AA421473
Em:AA326081 Em:H19142 Em:H29553 Em:AA345131 Em:H47026
Em:A1572262 Em:AA876086 Em:A1364862 Em:A1367012 Em:H92616
Em:H41010"
/evidence="not_experimental"
/product="DJ1104E15.1 (mannosyl (beta-1,4)-glycoprotein
beta-1,4-N-acetylglucosaminyltransferase)"
26520. .32881
/gene="MGAT3"
26834. .27125
/note="AluSg repeat: matches 7. .298 of consensus"
27188. .27489
/note="AluJb repeat: matches 1. .302 of consensus"
28035. .30123
/gene="MGAT3"

alignment_scores:
Quality: 261.00 Length: 340
Ratio: 1.442 Gaps: 19
Percent Similarity: 53.235 Percent Identity: 30.882

alignment_block:
US-08-656-811A-1 x HS1104E15 ..

Align seg 1/1 to: HS1104E15 from: 1 to: 112460

55 LeuGluProPheGluGluAspValLeu...GlyAlaGluTrpMet...G1 69
||||| ||| |||||: |||: |||: |||: |||: |||: |||
62446 CTCAAATGTTTGCAGAGGATGCTCTCCGGGACAGATTGGATCTTGA 62495
||: ||||| :||| |||||
62496 GAAATGGATTGAAGAGATTGCTGATGCC..... 62529
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
85 luArgLeuHisProPheGluSerAsnLeuGluPheThrSerLeuIle 101
||||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62530 .....CTGTTGGGTATAGATGACCTGGAA 62553

102 Thr...ProAspAspSerThrValSerLysAsp..... 111
||| ||||| |||: |||: |||: |||: |||: |||: |||
62554 ACCATGCCAGATGACCTTCTGACCACGTTGGATGACACTTCTGATCTCTT 62603
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112 .....lleLeuSerSerThrLeuGlnPheProThrGlnProValAsnI 126
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
62604 TGCCCCCTAGTCCAGGAGACTAATAAGCAGCCCCCAGACGGTGAAC. 62652
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
126 leProLeuTyAlaSerHisGlyAlaGluAspPheSerAlaGluThrGlu 142
|||||
62653 ..CAAATT..... 62658

143 PheGluAsnHisLeuSerProProAspSer.....ProGluGlnVa 156
||| ||||| |||: |||: |||: |||: |||: |||: |||
62659 .....GCCATCTC.....CCAGAAAGTTTAAACAAACCCGACGAGT 62696
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
156 lAlaProValIleAsnLeuGluProValGluLeuThrAlaSerHisMet 173
||||| |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||

```

[illegible]

seq_name: qb_ro:MUSTRANFAC

seq_documentation_block:	
LOCUS	MUSTRANFAC 1224 bp mRNA ROD 23-JUL-1993
DEFINITION	Mus musculus (clone Chopil) C/ATF mRNA, complete cds.
ACCESSION	L13791
VERSION	L13791.1 GI:293841
KEYWORDS	C/ATF; DNA-binding protein; transcription factor.
SOURCE	Mus musculus (library: lambda-ZAP of B. Spiegelman) cDNA to mRNA.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Vallejo,M., Ron,D., Miller,C.P. and Habener,J.F.
TITLE	C/ATF, a member of the activating transcription factor family of DNA-binding proteins, dimerizes with CAAT/enhancer-binding proteins and directs their binding to CAMP response elements
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 90, 4679-4683 (1993)
MEDLINE	93281642


```
429 CTCGGAATGGCCGGCTATGGATGATGGCTTGGCCAGTGCCTCAGACACCG 478
74 lyserPheLeuAspAlaLeuGlyAspAsnHisGluArgLeuHisProPhe 90
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
479 GCAAG...GAGGATGCCCTTCCGGGACAGATTGATGTTGGAGAAATG 525
91 GluSerAsnLeuLeuGluPheThrSerLeuLeuThrProAspSerTh 107
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
526 GATCTGAAAGAGTTTGATCTCGATCTCTGTTTCAATGGATGACCTGGA 575
107 rValSerLysAspLeuLeuSerSerThrLeu..... 117
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
576 AACCATGCCAGATGAGCTCTTGACACCATGTTGATGACACATGTGTATCTTT 625
118 .....GlnPheProThrGlnProValAsn 125
626 TTGCCCTCTAGTCCACAGACTATAGGAGCCCTCAGACAGTGAAC 675
126 IleProLeuTyAlaSerHisGlyAlaGluAspPheSerAlaGluThrG1 142
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
676 ...CCAATT..... 681
142 upheGluAsnHisLeuSerProProAspSer.....ProGluGlnV 156
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
682 .....GGCCATCTC.....CCAGAAAGTTTAAATAAAAGTCGACACAGG 718
156 aAlaProValIleAsnLeuGluPro.....ValGluLeu..... 164
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
719 TTGCCCTTTACATCTTGAGCCCTTTCCCTGTTCCCGAGGGTTCTG 768
165 .....ValGluLeu..... 167
769 TCTTCCACTCCAGACATTCCTTTAGTTAGAGCTAGGAGTGAAGTTGA 818
168 .....ThrAlaSerHisMetThrValI 175
819 TATCTCTGAAGAGACAGACACCTCTGCTCTGCTTACATTTACTCTAA 868
175 leSerProAspGlyLeuLeuGlyMetGluLeuAlaSerGluSerLeu 191
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
869 TCCCTCCA..... 876
192 ThrPheThrGluLeuAspPheValAsnPheAsnAspSerAlaValGlyse 208
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
877 TGTGTAAAGAGGAAGACACCTCTGCTGACATGACAGT.....GGCAT 920
208 rIleGlyAlaGluLeuLeuGlySerProLeuSerValAspAspV 225
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
921 CTGTATGAGCCCGGAGTCTACCTGGGCTCTCCC..... 954
225 alGluSerThrIleSerPheSerGlyProSerSerProGluThrSerGln 241
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
955 .....CAGCATAGCCCTCCCTCCACCTCCAGG 978
242 SerSerIleIleGluSerSerProGluLeuTyLysValIleSerThrse 258
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
979 .....GCCCCACCAGAC.....AATCTGCTCTCTCCAGG 1007
258 rSerIleAspAlaSerLysArgPheSerProTySerArgSerSerLysS 275
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
1008 TGGTTCGGTGGGCTCTCTCGGCCCAACCTTATGAC..... 1044
275 erLysGlnSerValLysThrSerAspAlaLysAlaProArgLysThrArg 291
1044 ..... 1044
292 ThrProAlaGlnProValProGluHisValIleMetGluHisLeuAspLys 308
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
1045 CCACCTGGAGTTAGTTTGCACAGCTAAAGTGAAGACTGAGAAATGGATAA 1094
308 sLysAspArgLysLysLeuGlnAsnLysAsnAlaIleArgTyArgM 325
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
1095 GAAGCTGAAAGATGGAGCAACACAGACAGCAGCAGCTAGGTACCGCC 1144
```

```
325 eTLYSLYSGLYGLUAlaGlnGlyIleLysGlyGluGlnGluLeu 341
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
1145 AGAAGAAGGGGCTGAGCAGAGGCCCTCAGTGGCGAGTGAAGGAGCTA 1194
342 GluGluLeuAsnThrLysLeuLysThrLysValAspAspLeuGlnArgG1 358
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
1195 GAAAAAAGAATGAGGCTCTGAAGAGAGAGGAGAGATTCTCTGGCCCAAGGA 1244
358 uIleLysTyrrMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
1245 GATCCAGTATCTGAAGACCTGATAGAAGAGTCGTAAGCAAGGGG 1293
seq_name: gb_ro:MMATF4
seq_documentation_block:
LOCUS MMATF4 1365 bp mRNA ROD 25-NOV-1991
DEFINITION M.musculus ATF-4 gene for activating transcription factor 4.
ACCESSION X61507
VERSION X61507.1 GI:50049
KEYWORDS activating factor; activating transcription factor; ATF-4 gene;
nuclear protein; transcription factor.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1365)
AUTHORS Pruitt,S.C.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-1991) S.C. Pruitt, Roswell Park Cancer Institute,
Dept. of Molecular and Cellular Biology, C & V 501, Elm and Carlton
Streets, Buffalo, NY 14263, USA
REFERENCE 2 (bases 1 to 1365)
AUTHORS Mielnicki,L.M. and Pruitt,S.C.
TITLE Isolation and nucleotide sequence of a murine cDNA homologous to
human activating transcription factor 4
JOURNAL Nucleic Acids Res. 19 (22), 6332 (1991)
MEDLINE 92066493
FEATURES
source
1..1365 Location/Qualifiers
/organism="Mus musculus"
/strain="129/sv"
/db_xref="taxon:10090"
/cell_type="embryonal sarcoma (EC) stem cell"
/cell_line="p9"
/clone_lib="plasmid; pcdPolyBN-"
/clone="pcdmATF4"
176..1321
/gene="ATF-4"
176..1321
/gene="ATF-4"
/codon_start=1
/product="activating transcriptionn factor 4"
/protein_id="CAA43723.1"
/db_xref="GI:50050"
/db_xref="MGD:MGI:88096"
/db_xref="SPTREMBL:O61328"
/translation="MALFTKSSSVAVTDKTFELSTFLESSKAPQHDDELPEQRSV
GGGLDDFFDQSGLGAEISGLDDYLEVAHKLHPGFSSDKAGSSSEWPMDDGLASAS
DTKEDAFSGTDHMLEKMDLKEFDALFRMDLETPDLELLTTLDDTCDLFLAPLVOE
TKNEPQTVPNTGHLPELTKVDQVAPPTFLQPPCGVGLSSTPHSFSLELGSVD
ISEGDRKPDSAAVITLIPPCVKEDTSPSDNGICMSPEYLGSPQHSPTSTRAPPDN
LPSGGSGSRGPRKPYDPGVSILTAKEVTEKLDKLKKMEQNTATNRYRQKRAEQE
ALTGEKELEKKNALKEKADSLAKEIQILKLDIEVRKARKKRP"
317..385
/gene="ATF-4"
/notes="acidic domain 1"
446..646
/gene="ATF-4"
/notes="acidic domain 2"
1115..1267
/gene="ATF-4"
/notes="DNA binding/dimerization domain (bz1p)"
misc_feature
317..385
misc_feature
446..646
misc_feature
1115..1267
```

BASE COUNT 334 a 354 c 368 g 309 t
ORIGIN

alignment_scores:
Quality: 255.00 Length: 434
Ratio: 1.288 Gaps: 21
Percent Similarity: 45.622 Percent Identity: 27.880

alignment_block:

US-08-656-811a-1 x MMATF4 ..

Align seg 1/1 to: MMATF4 from: 1 to: 1365

```

13 AlaArgGluTrpGlyLeuGluMetProValValGlnThrAspGlyGlnPh 29
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
240 GCACATTCTGGAATCCAGCAAGCCCAACACATGACCGAGATGAGCTT 289
29 e.....GlyAspLeu.....LysSerThrSerA 37
|.....|.....|.....|.....|.....|.....|.....|.....|
290 CCTGAACACGCGAAGTGTGGCGGGGACTTGATGACCCCTTCGACCAATC 339
||.....|.....|.....|.....|.....|.....|.....|.....|
340 CGGTTGGGGCTGAAGAAAGCCTAGGTCTCTTAGATGACTATCTGGAGG 389
||.....|.....|.....|.....|.....|.....|.....|.....|
50 ..GlyAlaThrLeuLysLeuGluProPheGluGluAspValLeuGly... 64
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
390 TGGCCAAGCACTTGAACCTCATGTTGCTTCCAGGCACAAAGCGGCTCC 439
65 AlaGluTrp.....MetGluSerSerAspLeuG1 74
||||:||||:||||:||||:||||:||||:||||:||||:||||:
440 TCGGAATGGCGGCTATGGATGATGGCTTGCCAGTGCCTCAGACACCGG 489
74 ySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeuHisProPheG 91
|:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
490 CNAAG..GAGATGCTCTTCGGGACAGATGGATGTGGAGAAATGG 536
91 LuSerAsnLeuGluPheThrSerLeuLeuThrProAspAspSerThr 107
||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
537 ATCTGAAAGAGTTTCACTTCGATGCTCTGTTTTCGAATGGATGCACCTG 586
108 ValSerLysAspIleLeuSerSerThrLeu..... 117
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
587 ACCATGCCAGATGACCTCTTCACCACTGTTGGATGACACATGTGATCTTT 636
118 .....GlnPheProThrGlnProValAsnI 126
637 TCCCTCTAGTCCAGAGACTAATAAGGAGCCCTCAGACAGTGAAC. 685
126 leProLeuTyraLaserHisGlyAlaGluAspPheSerAlaGluThrGlu 142
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
686 ..CCAATT..... 691
143 PheGluAsnHisLeuSerProProAspSer.....ProGluGlnVa 156
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
692 .....GCCCATCTC.....CCAGAAAGTTAATAAAAGTCGACCAAGT 729
156 lAlaProValIleAsnLeuPro..... 164
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
730 TGCCCCCTTACATCTTGCAGCCCTTCCCTGTTCCCGAGGGGTCTGT 779
165 .....ValGluLeu..... 167
780 CTTCCACTCCAGAGCATCTTTAGTTTAGAGCTAGGAGTCAAGTTGAT 829
168 .....ThrAlaSerHisMetThrValI 175
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
830 ATCTCTGAAGGAGACAGGAAGCCCTGACTCTGCTGCTTACATTACTCTAAT 879
175 eSerProAspGlyLeuLeuGlyGlyMetGluLeuAlaSerGluSerLeuT 192
|.....|.....|.....|.....|.....|.....|.....|.....|
880 CCTCCA.....T 887

```

```

192 hrPheThrGluLeuAspPheValAsnPheAsnAspSerAlaValGlySer 208
||| ||| ::| ||||| |||
888 GTGTAAAGGAGGAGACACTCCCTCTGTGACATGACAGT.....GCCATC 931
209 IleGlyGlyAlaGluGluLeuLeuGlySerProLeuSerValAspVa 225
::: ||||: ||||| |||||
932 TGTATGACCCGGAGTCTTACCTGGGCTCTCCC..... 964
225 lGluSerThrIleSerPheSerGlyProSerSerProGluThrSerGlnS 242
965 .....CAGCATAGCCCTCCACCTCCAGG. 988
242 erSerIleGluSerSerProGluLeuTyrlLysValIleSerThrSer 258
::: ||||: ::| ||| |||||
989 .....GCCCCACCAGAC.....AATCTGCCTTCTCCAGGT 1018
259 SerIleAspAlaSerLysArgPheSerProTyrlSerArgSerSerLysSe 275
::: |||| ||| ::||| ||||:
1019 GGTTCCTGGGTCTCTCGGCCCAACCTTATGAC..... 1054
275 rLysGlnSerValLysThrSerAspAlaLysAlaProArgLysThrArgT 292
1055 .....C 1055
292 hrProAlaGlnProValProGluHisValIleMetGluHisLeuAspLys 308
||||: ||| ||| |||||
1056 CACCTGGAGTTAGTTTGACAGCTAAAGTCAAGACTCAGAAATGGATAAG 1105
309 LysAspArgLysLysLeuGlnAsnLysAsnAlaAlaIleArgTyrlArgMe 325
||| ||||| ||||| ||||| ||||| ||||| |||||
1106 AAGCTGNAAGATGGAGCAAAACAGACAGCAGCCACTAGGTACCGCA 1155
325 tLysLysLysGlyGluAlaGlnGlyIleLysGlyGluGlnGluLeuG 342
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1156 GAAGAGCGGGCTGAGCAGGAGGCCCTCACTGGCAGTGTAGGAGCTAG 1205
342 luGluLeuAsnThrLysLysLysThrLysValAspAspLeuGlnArgGlu 358
|||: ||| ||||| ||||:||||:||||:||||:
1206 AAAAAAGAAATCAGGCTCTGAAAGAGAGGAGGAGATTCCTCGCCCAAGGAG 1255
359 IleLysTyrlMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1256 ATCCAGTATCTGAAAGACCTGTAGAGAGGTCCTGTAAGGCAAGGGG 1303

```

seq_name: gb_ov:AB013138

seq_documentation_block:

LOCUS AB013138 1461 bp mRNA VRT 02-APR-1999
DEFINITION Gallus gallus mRNA for ATP4, complete cds.

ACCESSION AB013138

VERSION AB013138.1 GI:4586686

KEYWORDS ATP4.

SOURCE Gallus gallus 3.5 day old eye cDNA to mRNA.

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;

Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (sites)

Murata,T., Suzuki,H., Ogino,H. and Yasuda,K.

Isolation and analysis of embryonic expression pattern of chicken

ATF4

JOURNAL Unpublished (1998)

REFERENCE 2 (bases 1 to 1461)

AUTHORS Murata,T.

TITLE Direct Submission

JOURNAL Submitted (21-APR-1998) to the DDBJ/EMBL/GenBank databases. Takuya

Murata, Nara Institute of Science and Technology, Graduated

School of Biological Science; 8916-5 Takayama-cho, Ikoma, Nara

630-0101, Japan (E-mail:t-murata@bs.ais-t-nara.ac.jp,

Tel.:81-743-72-5550, Fax:+81-743-72-5559)

FEATURES

source

1. .1461

/organism="Gallus gallus"

Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S.,
Schageman, J., Schultz, R.A., Stimson, S., Syed, M. and Ward, T.

TITLE
JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 80920)
Evans, G.A., Bradbury, P., Brignac, S., Bumeister, R., Burbee, D.,
Davie, J., Davies, C.J., Davis, C., English, C., Fondon, T.,
Franklin, T.L., Garner, H.R., Gordon, M., Gotway, G., Grant, O.,
Hahner, L., Harris, J., Hinson, S., Megarity, C., Narayanaswamy, U.,
Newton, J., O'Brien, K., Oliver, T., Patel, P., Probst, S., Rayner, S.,
Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D.,
Ward, T. and Wilson, R.

TITLE
JOURNAL

Direct Submission
Submitted (24-APR-1997) Genome Science and Technology Center,
University of Texas Southwestern Medical Center at Dallas, 5323
Harry Hines Blvd, Dallas, TX 75235-8591, USA

REFERENCE
AUTHORS

3 (bases 1 to 80920)
Evans, G.A., Athanasios, M., Basit, M., Bradbury, P., Brignac, S.,
Bumeister, R., Davis, C., English, C., Franklin, T.L., Garner, H.R.,
Gee, V., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J.,
Hinson, S., Narayanaswamy, U., Newton, J., O'Brien, K., Patel, P.,
Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D.,
Ward, T. and Wilson, R.

TITLE
JOURNAL

Direct Submission
Submitted (17-DEC-1997) Genome Science & Technology Center,
University of Texas Southwestern Medical Center, 5323 Harry Hines
Blvd, Dallas, TX 75235-8591, USA

REFERENCE
AUTHORS

4 (bases 1 to 80920)
Evans, G.A., Athanasios, M., Aguayo, P., Armstrong, D., Basit, M.,
Buettnner, J., Butler, C., Card, P., deSailboat, F., Dunn, J.,
English, C., Ethridge, S., Garner, H.R., Gee, V., Gordon, M., Gotway, G.,
Grant, O., Hahner, L., Joslin, J., Lewis, E., Loo, H., Loo, K.N.,
Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S.,
Schageman, J., Schultz, R.A., Stimson, S., Waller, K. and Ward, T.

TITLE
JOURNAL

Direct Submission
Submitted (30-OCT-1998) Genome Science & Technology Center,
University of Texas Southwestern Medical Center, 5323 Harry Hines
Blvd, Dallas, TX 75235-8591, USA

COMMENT

On Oct 30, 1998 this sequence version replaced gi:2695562.
IMPORTANT: This submission contains the entire insert of clone
pDJ356d6. pDJ356d6 comes from the RPCI-3 PAC library constructed
at the Roswell Park Cancer Institute by the Pieter de Jong group.
This clone has been finished according to strict quality criteria
and attempts have been made to resolve all base calling problems
such as compressions and repetitive elements. The expected
Phred/Phrap calculated errors/10kb is 0.49. In addition, attempts
have been made to assure 99% of consensus base calls consist of
either double-stranded coverage or 2 types of labeling chemistry on
one strand.

FEATURES
source

1. .80920
/organism="Homo sapiens"
/db_xref="taxon:9606"
complement(2..356)
/rpt_family="Alu"
repeat_region
complement(437..723)
/rpt_family="Alu"
repeat_region
1892..2123
/rpt_family="Alu"
repeat_region
complement(2426..3043)
/rpt_family="L1"
repeat_region
complement(3085..3379)
/rpt_family="Alu"
repeat_region
4272..4528
/rpt_family="Alu"
repeat_region
4543..5068
/rpt_family="MLT2B2"
repeat_region
complement(5137..6056)
/rpt_family="Alu"
repeat_region
complement(6085..6628)
/rpt_family="L1"
repeat_region
complement(7044..7319)
/rpt_family="Alu"
repeat_region

complement(8255..8853)
/rpt_family="Alu"
repeat_region
10496..10773
/rpt_family="Alu"
repeat_region
11241..11526
/rpt_family="Alu"
repeat_region
12510..12658
/rpt_family="Alu"
repeat_region
12822..12961
/rpt_family="MIR"
repeat_region
complement(13481..13750)
/rpt_family="Alu"
repeat_region
14712..14951
/rpt_family="Alu"
repeat_region
complement(15508..15827)
/rpt_family="Alu"
repeat_region
complement(16084..16373)
/rpt_family="Alu"
repeat_region
17076..17373
/rpt_family="Alu"
repeat_region
complement(17603..17879)
/rpt_family="Alu"
repeat_region
complement(22409..22688)
/rpt_family="Alu"
repeat_region
23561..23821
/rpt_family="Alu"
repeat_region
complement(23962..24258)
/rpt_family="Alu"
repeat_region
complement(24276..24577)
/rpt_family="Alu"
repeat_region
complement(24608..24960)
/rpt_family="L1"
repeat_region
complement(24947..25248)
/rpt_family="Alu"
repeat_region
complement(25412..25623)
/rpt_family="L1"
repeat_region
complement(25719..26007)
/rpt_family="Alu"
repeat_region
complement(26182..26449)
/rpt_family="Alu"
repeat_region
complement(26809..27084)
/rpt_family="Alu"
repeat_region
complement(29370..30095)
/rpt_family="Alu"
repeat_region
complement(33313..33601)
/rpt_family="Alu"
repeat_region
34701..34813
/rpt_family="Alu"
repeat_region
36390..36629
/rpt_family="Alu"
repeat_region
37020..37297
/rpt_family="Alu"
repeat_region
complement(39560..39836)
/rpt_family="Alu"
repeat_region
complement(42468..42779)
/rpt_family="Alu"
repeat_region
42805..43054
/rpt_family="Alu"
repeat_region
complement(43105..43388)
/rpt_family="Alu"
repeat_region
complement(44130..44313)
/rpt_family="L1"
repeat_region
complement(44453..44745)
/rpt_family="Alu"
repeat_region
44794..45070
/rpt_family="Alu"
repeat_region
46814..46880
/rpt_family="Alu"
repeat_region
complement(46943..47217)
/rpt_family="Alu"
repeat_region
49439..49641
/rpt_family="Alu"
repeat_region
49999..50284
/rpt_family="Alu"
repeat_region


```
seq_name: gb_ro:AB012277

seq_documentation_block:
LOCUS      AB012277      880 bp      mRNA      ROD      25-AUG-1999
DEFINITION Mus musculus mRNA for ATF4, partial cds.
ACCESSION  AB012277
VERSION    AB012277.1 GI:2988374
KEYWORDS   ATF4.
SOURCE     Mus musculus cDNA to mRNA.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS    Nishizawa, M. and Nagata, S.
TITLE      cDNA clones encoding leucine-zipper proteins which interact with
            G-CSF gene promoter element 1-binding protein
JOURNAL    FEBS Lett. 299 (1), 36-38 (1992)
MEDLINE    92183900
REFERENCE  2 (bases 1 to 880)
            Nishizawa, M. and Nagata, S.
            Direct Submission
            Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Mikio
            Nishizawa, Kansai Medical University, Department of Medical
            Chemistry; 10-15 Fumizono, Moriguchi, Osaka 570-8506, Japan
            (E-mail:nishizawa@takii.kmu.ac.jp, Tel:81-6-992-1001(ex.2454),
            Fax:81-6-992-1781)

FEATURES             Location/Qualifiers
     source           1..880
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS              <1..792
                     /product="ATF4"
                     /protein_id="BAA25314.1"
                     /db_xref="GI:2988375"
                     /translation="KMDLKEFFDALFRMDLTMPDELLTTLDDTCDLFPAPLVQETN
                     KEPQVNPVIGHLPESLIRKVDQVAPTFIQPFCSFGVLSSTPEHSFSLGLSEVDLS
                     EGGKPDSPRAITLIPPCVKEDTSDNSGSCMSPEYLGSPSPSTRAPPDNLIP
                     SPGSGSRSPKPDYDPGVSLTAKVTEKLDKLRKMEQNKTAATRYRQKRAREQAL
                     TGECKELEKKEALKEKSLAKEIQYLKDLIEVRKARKGRKP"
     polyA_signal     849..854
     BASE COUNT      246 a 207 c 218 g 209 t
     ORIGIN
1       polyA_signal
2       BASE COUNT
3       ORIGIN

alignment_scores:
Quality: 222.00      Length: 328
Ratio: 1.563         Gaps: 13
Percent Similarity: 43.293 Percent Identity: 27.439

alignment_block:
US-08-656-811a-1 x AB012277 ..
Align seg 1/1 to: AB012277 from: 1 to: 880

96  GluPheThrSerLeuIleThProAspAspSerThrValSerLysAspTl 112
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
22  GACTTCGATGCTCTCTTTTTCGAATGGATGACCTGGAAACCATGCCAGATGA 71
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
112 eLeuSerSerThrLeu.....
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
72  GCTCTTGACACAGGTGGATGACACATGATCTTTTTCCTTCCTAGTCC 121
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
118 .....GlnPheProThrGlnProValAsnIleProLeuTyrAla 130
    :||| :||| :||| :||| :||| :||| :||| :|||
122  AAGACACTAATAAGAGGCCCTCAGACAGTGAAC...CCAAATT..... 162
    :||| :||| :||| :||| :||| :||| :||| :|||
131  SerHisGlyAlaGluAspPheSerAlaGluThrGluPheGluAsnHisLe 147
    :||| :||| :||| :||| :||| :||| :||| :|||
163 .....GGCCATCT 170

147  userProProAspSer.....ProGluGlnValAlaProValIleA 161
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
```

```
171  C.....CCAGAAAGTTTAAATAAAGTCGACAGGTTGCCCTTTTACAT 214
161  snLeuGluPro.....
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
215  TCTTGACAGCTTTCCCTCTGTTCCCGAGGGTCTGTCTTCCACTCCAGAG 264
165  .....ValGluLeu.....
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
265  CATTCCTTTAGTTTAGAGCTAGGCAGTGAAGTTGATATCTCTGAAGGAGA 314
168  .....ThrAlaSerHisMetThrValIleSerProAspGlyL 180
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
315  CAGGAGAGCTGACTCTGCTGCTTACTTACTTAATCCCTCCA..... 357
180  euLeuGlyGlyMetGluLeuAlaSerGluSerLeuThrPheThrGluLeu 196
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
358  .....TGCTGTAAGAGGAGAA 372
197  AspPheValAsnPheAsnAspSerAlaValGlySerIleGlyAlaG 213
    :||| :||| :||| :||| :||| :||| :||| :|||
373  GACACTCCCTCTGCACAAATGACAGT.....GGCATCTGTATGAGCCCGA 416
213  uGluLeuLeuGlySerProLeuSerValAspAspValGluSerThrIleS 230
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
417  GTCTTACCTGGGCTCTCCC..... 435
230  erPheSerGlyProSerProGluThrSerGlnSerSerIleIleGlu 246
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
436  .....CAGCATAGCCCTCCACCTCCAGG..... 459
247  SerSerProGluLeuTyrLysValIleSerThrSerSerIleAspAla 263
    :||| :||| :||| :||| :||| :||| :||| :|||
460  GCCCCACACAGAC.....AATCTGCTTCTCCAGGTGGTCCCGTGGGTC 503
263  rlyArgPheSerProTyrSerArgSerSerLysSerLysGlnSerVal 280
    :||| :||| :||| :||| :||| :||| :||| :|||
504  TCCTCGGCCCAACCTTATGAC..... 525
280  ysThrSerAspAlaLysAlaProArgLysThrArgThrProAlaGlnPro 296
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
526  .....CCACCTGGAGTTAGT 540
297  ValProGluHisValIleMetGluHisLeuAspLysLysAspArgLys 313
    :||| :||| :||| :||| :||| :||| :||| :|||
541  TTGCACAGCTAAAGTCAAGACTGAGAAATTTGATAAGAGCTGAAAAAGAT 590
313  sLeuGlnAsnLysAsnAlaAlaIleArgTyrArgMetLysLysLysGly 330
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
591  GGAGCAAAACAAGACAGACAGCCACTAGGTACCCCCAGAGAAGCGGGCTG 640
330  lAlaGlnGlyIleLysGlyGluGluGlnGluLeuGluLeuAsnThr 346
    :||| :||| :||| :||| :||| :||| :||| :|||
641  AGCAGAGGCCCTCACTGCGGAGTGTAAAGAGCTAGAAAAAAGATGAG 690
347  LysLeuLysThrLysValAspAspLeuGlnArgGluIleLysTyrMet 363
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
691  GCTCTGAAAGAGAAGCAGATTTCTCTGCGCAAGGAGATCCAGTATCTGAA 740
363  sAsnLeuMetGluAspValCysLysAlaLysGly 374
    :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
741  AGACCTGTAGAGAGGTCCTCGTAAGCAAGGGGG 774
```

```
seq_name: gb_pr2:HSU03712
seq_documentation_block:
LOCUS      HSU03712      1442 bp      DNA      PRI      20-MAY-1995
DEFINITION Human TAXREB67 pseudogene, complete sequence.
ACCESSION  U03712 L15294 L15295
VERSION    U03712.1 GI:434667
KEYWORDS   human.
SOURCE     human.
            ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

REFERENCE 1 (bases 1 to 1442)
 AUTHORS Das.S. and Gitschler.J.
 TITLE Identification of a TXREB pseudogene (TXREBP) located between the
 JOURNAL genes for p55 (MP1) and G6PD on Xq28
 MEDLINE Genomics 21 (1), 275-278 (1994)
 REFERENCE 94375031
 AUTHORS Das.S.
 TITLE Direct Submission
 JOURNAL Submitted (24-NOV-1993) Soma Das, Medicine, Howard Hughes Medical
 Institute, University of California, 3rd and Parnassus Avenues, San
 Francisco, CA 94143-0724, USA
 COMMENT On Dec 14, 1993 this sequence version replaced gi:292867.
 FEATURES
 Location/Qualifiers
 1..1442
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="YAC527(85-1963)"
 /chromosome="X"
 /map="Xq28"
 /cell_line="CGM-1"
 283..1363
 /note="pseudogene for DNA binding protein TAXREB67"
 /pseudo

CDS
 283..1363
 /note="pseudogene for DNA binding protein TAXREB67"
 /pseudo

BASE COUNT 350 a 375 c 396 g 321 t
 ORIGIN

alignment_scores:
 Quality: 220.00 Length: 407
 Ratio: 1.068 Gaps: 19
 Percent Similarity: 50.614 Percent Identity: 25.307

alignment_block:
 US-08-656-811a-1 x HS003712 ..
 Align seg 1/1 to: HS003712 from: 1 to: 1442

```

16 TrpGlyLeuGluMetProValValGlnThrAsp..... 26
|||||..... 19
320 TGGGGCTTGTATGCTCCCTTGACGAGTGGGCTGAGAAAG 369
|||||..... 19
27 ....GlyGlnPheGlyAspLeuLysSerThrSerArgHis..... 38
|||..... 19
370 CCTAGGCTCTTAGTACTACCTGGAGGTGGCCAGCACTTCAACCTC 419
|||..... 19
39 .....GlyGlyAspGluSerLeuSerLeuGln 47
|||||..... 19
420 ATGGGTTCTCCAGCGACAAGGCTAAGGCAGACTCCTCGAGTGGCTGGCT 469
|||||..... 19
48 ProGlnGlyAlaThrLeuLysLeuGluProPheGluGluAspVal...Le 63
|||||..... 19
470 GTGGATGGGTTGGTTCAGTGCTCCCAAGCATGGCAAGGAGGATGCTTCTT 519
|||||..... 19
63 uGlyAlaGluTrpMet...GluSerSerAspLeuGlySerPheLeuAspA 79
|||||..... 19
520 TGGACAGATTGGATGTTGGAGAAACTGATCTGAGAGTTC..... 562
|||||..... 19
79 laLeuGlyAspAsnHisGluArgLeuHisProPheGluSerAsnLeuLeu 95
|||||..... 19
562 ..... 562
96 GluPheThrSerLeuLeuThrProAspAspSerThrValSerLysAspI1 112
|||||..... 19
563 GACTTTGATGCCCTTGGTGGTATAGATGACCTGGAAACCATGCCAGGCA 612
|||||..... 19
112 eLeuSerSerThrLeuGlnPheProThrGlnProValAsnIleProLeuT 129
|||||..... 19
613 GCTTCTGCCACGTTGGAT..... 631
|||||..... 19
129 yrAlaSerHisGlyAlaGluAspPheSerAlaGluThrGluPheGluAsn 145
|||..... 19

```

```

632 .....GACTCGTGTGATCTCTTTGCCCCCTAGTCCAGGAGACT 670
146 HisLeuSerProAspSer.....ProGluG1 155
|||||..... 19
671 ATTAAGGAGCCCCCAGGAGTCCAGGAGCTATTAGGAGCCCCCAGAGAT 720
|||||..... 19
155 nValAlaProValIleAsnLeuGluProValGluLeuThr.....Ala. 169
|||||..... 19
721 GGTGAACCCGATTGGCCATCTC...CCAGAAAGTTTAACCCGACCAAGTT 767
|||||..... 19
170 SerHisMetThrValIleSerProAspGlyLeuGlyMetGluLe 186
|||||..... 19
768 GCCCCTTCACCTCTTACAACTCTTCCCTTCCAGGGTCCAGTC 817
|||||..... 19
186 uAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsnPheAsnA 203
|||||..... 19
818 CTCACCTCCAGATTATTCCTTTAGT...CTAGAG..... 848
|||||..... 19
203 spSerAlaValGlySerIleGlyAlaGluLeuLeuGlySerPro 219
|||||..... 19
849 .....CTGGCAGTGAA 860
|||||..... 19
220 LeuSerValAspAspValGluSerThrIleSerPheSerGlyProSerSe 236
|||||..... 19
861 GTGGATATC.....TTTGAAGGAGCTAGGAA 886
|||||..... 19
236 rProGluThrSerGlnSerSerIleIleGluSerProGluLeuTrL 253
|||||..... 19
887 GCCAGACTCC.....ACTGTTACATTTCCAGATCCTCCTGCACAA 930
|||||..... 19
253 ys.....ValIleSerThr 257
|||||..... 19
931 AGGAGGAGGAGCGCCCTCAGATAATGATAGTGGCATCTGTATGAGCCA 980
|||||..... 19
258 SerSerIleAspAlaSerLysArgPheSerProTyrSerArgSerSerLy 274
|||||..... 19
981 GAGTCCTATCTGGGTCTCCTCAGACAGCCCTCTACCAAGGGCTCTCC 1030
|||||..... 19
274 sSerLysGln.....SerValLysThrSerAspAlaLysAlap 287
|||||..... 19
1031 AAATAGGAGCTCCATCTCCAGGTGTTCTCTGTGGTCTGCCGT...C 1077
|||||..... 19
287 roArgLysThrArgThrProAlaGlnPro...ValProGluHisValIle 302
|||||..... 19
1078 CCAAACTTATGACCTCTCGGAGAGAAGATGGTAGCAGCACAAAGTAAG 1127
|||||..... 19
303 MetGluHisLeuAspLysLysAspArgLysLysLeu...GlnAsnLysAsn 318
|||||..... 19
1128 GGTGAGAAACTGGATAAGAGCTGAAAAAATAATGGACAAACCAACA 1177
|||||..... 19
319 AlaAlaIleArgTyrArgMetLysLysLysGlyGluAlaGlnGlyIleLy 335
|||||..... 19
1178 GCAGAACTAGGTATCGCCAGAGAAGAGGCGGAGGAGGAGGCTCTCAC 1227
|||||..... 19
335 sGlyGluGlnGluLeuGluGluLeuAsnThrLysLeuLysThrLysV 352
|||||..... 19
1228 TGGTGAGTGCAAGCGGTGGGAAAAGAGAACAGAGGCTCTGCAAGAGAGG 1277
|||||..... 19
352 alAspAspLeuGlnArgGluIleLysTyrMetLysAsnLeuMetGluAsp 368
|||||..... 19
1278 CGGATTCCTCGCGGAGGAGATCCAGTACATGAAAGATTTCGATAGAAG 1327
|||||..... 19
369 ValCysLysAlaLysGly 374
|||||..... 19
1328 GTCGCAAGGCAAGGGG 1345
|||||..... 19

```

seq_name: gb_pr1:AB021663

seq_documentation_block:
 LOCUS AB021663 1506 bp mRNA PRI 03-JUN-1999
 DEFINITION Homo sapiens mRNA for leucine-zipper protein, complete cds.
 ACCESSION AB021663
 VERSION AB021663.1 GI:4996450

```

KEYWORDS   leucine-zipper protein.
SOURCE     Homo sapiens cDNA to mRNA.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1506)
AUTHORS   Kohroki,J. and Tanaka,K.
TITLE     cDNA clone encoding leucine-zipper protein
JOURNAL   Published Only in Database (1999) in press
REFERENCE  2 (bases 1 to 1506)
AUTHORS   Kohroki,J. and Tanaka,K.
TITLE     Direct Submission
JOURNAL   Submitted (21-DEC-1998) to the DDBJ/EMBL/GenBank databases. Junya
           Kohroki, Osaka University, Graduate School of Pharmaceutical
           Sciences, 1-6, Yamadaoka, Suita, Osaka 565-0871, Japan
           (E-mail:Kohroki@phs.osaka-u.ac.jp, Tel:81-6-879-8230,
           Fax:81-6-879-8234)
FEATURES   Location/Qualifiers
            source          1..1506
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"
            CDS             118..666
                           /codon_start=1
                           /product="leucine-zipper protein"
                           /protein_id="BA878477.1"
                           /db_xref="GI:4996451"
                           /translation="MASLKKLEQMEDFDLAPPPPPPLPPPPPLPPSPLPS
                           LPSPDLQPPLVLTDLALYCRNEAQEVGMPPPPPPPPPPPPPSRLAPYPH
                           PATTRGDRKKQKDNKSAALRYRQRKABGELEGECQGLEARNRELKRAESVERE
                           IOXVKDLLIEVYKRSQRTSC"
            polyA_site      1506
                           /note="13 a nucleotides"
            BASE COUNT     325 a 455 c 411 g 315 t
            ORIGIN
            alignment_scores:
              Quality: 188.50      Length: 146
              Ratio: 2.094         Gaps: 2
              Percent Similarity: 61.644 Percent Identity: 30.822
            alignment_block:
            US-08-656-811A-1 x AB021663 ..
            Align seg 1/1 to: AB021663 from: 1 to: 1506
            234 ProSerProGluThrSerGlnSerSerIleIleGluSerSerProG1 250
            ||||| ||| : : : ||| : : : : : : : : : : : : : : :
            232 CCTCCTCCCTCCCTGTCCTCCCTCCCTTGTGACCTCCGCCACGCCCTGT 281
            |||
            250 uleuTyLysValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
            ||| : : : : : : : : : : : : : : : : : : : : :
            282 CTTG.....GATACCTGGAGCTTGCTGG 304
            267 erProTySerArgSerSerLysSerLysGlnSerValLysThrSerAsp 283
            : : ||| : : : : : : : : : : : : : : : : : : : :
            305 CCATCTACTGCTCCGACACGAGCGCGGAGGAGGAGTGCGGCGCT 354
            284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGlu.. 299
            ||| : : : : : : : : : : : : : : : : : : : :
            355 CTGCCCGCCGACGAGCGCCCTCCTCTCTCCACCTCAACCTTCTCG 404
            300 .....HisValIleMetGluHisLeuAspLysLysAspA 311
            ||| : : : : : : : : : : : : : : : : : : : :
            405 CTGGCCCCCTACCCACATCTCTGCCACCACCGAGGGGACCCGACCA 454
            311 rGlyLysLeuGlnAsnLysAsnAlaAlaIleArgTyrArgMetLysLys 327
            : : ||| : : : : : : : : : : : : : : : : : : : :
            455 AGAAGAGACCAACAACAGTCCGGCGCTCTGAGGTACCGCCAGCGGAAG 504
            328 LysGlyGluAlaGlnGlyLysGlyGluGlnGlnGluLeu 344
            : : : : : : : : : : : : : : : : : : : :
            505 CGGGCAGAGGGTGGAGCCCTGGAGGCGGAGTCCAGGGGCTGGAGGCAG 554

```

```

344 uAsnThrLysLeuLysThrLysValAspAspLeuGlnArgGluIleLysT 361
||| : : : : : : : : : : : : : : : : : : : :
555 GAATCCGAGCTCAAGGAACGGCGAGTCCGTGGAGCGGAGATCCAGT 604
361 yrMetLysAsnLeuMetGluAspValCysLysAlaLys 373
||| : : : : : : : : : : : : : : : : : : : :
605 ACGTCAAGGACCTGCTCATCGAGGTTTACAGGCCCGG 642
seq_name: gb_htg24:AC011452
seq_documentation_block:
LOCUS   AC011452 147307 bp DNA HTG 06-APR-2000
DEFINITION Homo sapiens chromosome 19 clone CTC-326K19, WORKING DRAFT
SEQUENCE, 3 unchromed pieces.
ACCESSION AC011452
VERSION AC011452.3 GI:7458719
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
          ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          REFERENCE 1 (bases 1 to 147307)
          AUTHORS DOE Joint Genome Institute.
          TITLE Sequencing of Human Chromosome 19
          JOURNAL Unpublished
          REFERENCE 2 (bases 1 to 147307)
          AUTHORS DOE Joint Genome Institute.
          TITLE Direct Submission
          JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
          Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
          On Apr 6, 2000 this sequence version replaced gi:6604391.
          -----Genome Center
          Center: Joint Genome Institute
          Center Code: JGI
          Web site: http://www.jgi.doe.gov
          -----Summary Statistics
          Consensus quality: 145829 bases at least Q40
          Consensus quality: 146455 bases at least Q30
          Consensus quality: 146724 bases at least Q20
          Estimated insert size: 147307; sum-of-contigs estimation
          Estimated insert size: 147530; agarose-fp estimation
          Quality coverage: 6.40x in Q20 bases; agarose-fp estimation
          Quality coverage: 6.41x in Q20 bases; sum-of-contigs estimation
          -----
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 3 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          * 1 22842: contig of 22842 bp in length
          * gap of unknown length
          * 22843 81048: contig of 58206 bp in length
          * gap of unknown length
          * 81049 147307: contig of 66259 bp in length.
          FEATURES
            Location/Qualifiers
              1..147307
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="19"
              /clone="CTC-326K19"
            BASE COUNT 36775 a 37696 c 38022 g 34814 t
            ORIGIN
            alignment_scores:
              Quality: 188.50      Length: 146
              Ratio: 2.094         Gaps: 2
              Percent Similarity: 61.644 Percent Identity: 30.822

```

```
alignment_block:
US-08-656-811A-1 x AC011452  ..
Align seg 1/1 to: AC011452 from: 1 to: 147307

234 ProSerSerProGluThrSerGlnSerSerIleIleGluSerSerProG1 250
|||||  |||  ::  |||  ::  ::  |||
120204 CCTCCCTCCCTGTCCTCCCTCCCTTTGACCTCCGCCACGCCCTGT 120253
250 uLeuTyrLysValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
|||  |||  ::  |||  ::  ::  |||
120254 CTGG.....GATACTCTGGACTGTGCTGG 120276

267 erProTyrSerArgSerSerLysSerLysGlnSerValLysThrSerAsp 283
::  |||  |||  ::  ::  ::  |||
120277 CCATCTACTGCGCAACGAGCGCGGAGGAGGAAGTGGGGATGCCGCT 120326

284 AlalysAlaProArgLysThrArgThrProAlaGlnProValProGlu.. 299
|||  ::  |||  ::  |||  |||  |||
120327 CTGCCCCCGCCACAGCAGCCCTCTCTCTCCACCTCAACCTTCTCG 120376

300 .....HisValIleMetGluHisLeuAspLysLysAspA 311
|||  ::  |||  ::  |||  ::  |||
120377 CTGCCCCCTACCCACATCTGTCACACCCGAGGGGACCGCAGCAAA 120426

311 rGLysLysLeuGlnAsnLysAsnAlaIleArgTyrArgMetLysLys 327
::  |||  |||  |||  |||  |||  |||  |||  |||  |||
120427 AGAAGAGAGACAGACAAGTCGCGGCTCTGAGGTACCGCCAGCGGAAG 120476

328 LysGlyGluAlaGlnGlyIleLysGlyGluGlnGluLeuGluGluLe 344
::  |||  |||  |||  |||  |||  |||  |||  |||  |||
120477 CGGCAGAGGTGAGGCCCTGGAGGCGGAGTCCAGGGGCTGGAGGCACG 120526

344 uAsnThrLysLeuLysThrLysValAspLysLeuGlnArgGluIleLysT 361
|||  ::  |||  |||  ::  ::  |||  ::  |||  |||  |||
120527 GAATCGGAGCTGAGGAGAACCGGAGAGTCCGTGGAGCGGAGATCCAGT 120576

361 yrMetLysAsnLeuMetGluAspValCysLysAlaLys 373
|||  |||  |||  |||  |||  |||  |||  |||  |||
120577 ACGTCAAGGACCTGCTCATCGAGGTTTACAGGCCCGG 120614
```

OM of: US-08-656-811A-1 to: N_Geneseq_36.* out_format : pfs

Date: Aug 8, 2000 8:43 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame-p2n.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US08656811/runat_25072000_142318_17131/app_query.fasta_1.443
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=100000 -USER=US08656811_@Cgn1_1_88
-NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-656-811A-1

Query length: 379

Database: N_Geneseq_36.*

Database sequences: 311585

Database length: 125096042

Search time (sec): 54.050000

score_list:

Sequence	Strd Orig	ZScore	EScore Len	Documentation
N_Geneseq_36:V04079	+ 1923.00	2891.57	1.8e-153	1336 ! Aplysia cAMP-response element
N_Geneseq_36:T77784	+ 137.50	199.36	0.0017	669 ! cDNA clone 59 encoding NIP-1. N
N_Geneseq_36:V29204	+ 132.00	175.13	0.0373	3190 ! Gene encoding the mouse cAMP r
N_Geneseq_36:T14025	+ 128.50	183.96	0.0120	798 ! Drosophila dCREB1 cDNA. Novel c
N_Geneseq_36:V69717	+ 124.50	161.41	0.2169	4031 ! Tumour rejection antigen precu
N_Geneseq_36:V69720	+ 124.50	160.93	0.2307	4225 ! Tumour rejection antigen precu
N_Geneseq_36:T17502	+ 121.50	152.97	0.6404	5914 ! Mutated BRCA1 coding sequence
N_Geneseq_36:Q66090	+ 120.50	169.52	0.0767	1005 ! MUSCJUNX, a sample target spec
N_Geneseq_36:Q13674	+ 120.50	156.75	0.3944	3518 ! E2A.E12 transcript. Detection
N_Geneseq_36:T17471	+ 120.50	151.46	0.7775	5914 ! Mutated BRCA1 coding sequence
N_Geneseq_36:T17506	+ 120.50	151.46	0.7775	5914 ! Mutated BRCA1 coding sequence
N_Geneseq_36:V32452	+ 119.50	145.69	1.63	8982 ! Human receptor tyrosine kinase
N_Geneseq_36:T42021	+ 118.50	148.89	1.08	5656 ! BRCA1 allele #5803. New mutant
N_Geneseq_36:T70064	+ 118.50	148.89	1.08	5656 ! Mutant BRCA1 allele #5803. BR
N_Geneseq_36:V60560	+ 118.50	148.89	1.08	5656 ! Tumourigenic BRCA1 allele #580
N_Geneseq_36:T42023	+ 118.50	148.83	1.09	5689 ! BRCA1 allele #9815. New mutant
N_Geneseq_36:T70066	+ 118.50	148.83	1.09	5689 ! Mutant BRCA1 allele #9815. BR
N_Geneseq_36:V60562	+ 118.50	148.83	1.09	5689 ! Tumourigenic BRCA1 allele #981
N_Geneseq_36:T42031	+ 118.50	148.80	1.09	5707 ! BRCA1 allele #582. New mutant
N_Geneseq_36:T70074	+ 118.50	148.80	1.09	5707 ! Mutant BRCA1 allele #582. BR
N_Geneseq_36:V60570	+ 118.50	148.80	1.09	5707 ! Tumourigenic BRCA1 allele #582
N_Geneseq_36:T42027	+ 118.50	148.80	1.09	5709 ! BRCA1 allele #6401. New mutant
N_Geneseq_36:T42022	+ 118.50	148.80	1.09	5709 ! BRCA1 allele #9601. New mutant
N_Geneseq_36:T70065	+ 118.50	148.80	1.09	5709 ! Mutant BRCA1 allele #9601. BR
N_Geneseq_36:T70070	+ 118.50	148.80	1.09	5709 ! Mutant BRCA1 allele #6401. BR
N_Geneseq_36:V11738	+ 118.50	148.80	1.09	5709 ! Human BRCA1 DNA 3888delGA mut
N_Geneseq_36:V60561	+ 118.50	148.80	1.09	5709 ! Tumourigenic BRCA1 allele #961
N_Geneseq_36:V60566	+ 118.50	148.80	1.09	5709 ! Tumourigenic BRCA1 allele #64
N_Geneseq_36:T42026	+ 118.50	148.79	1.09	5710 ! BRCA1 allele #388. New mutant
N_Geneseq_36:T70069	+ 118.50	148.79	1.09	5710 ! Mutant BRCA1 allele #388. BR
N_Geneseq_36:V60565	+ 118.50	148.79	1.09	5710 ! Tumourigenic BRCA1 allele #388
N_Geneseq_36:T42024	+ 118.50	148.79	1.09	5711 ! BRCA1 allele #8403. New mutant
N_Geneseq_36:T70067	+ 118.50	148.79	1.09	5711 ! Mutant BRCA1 allele #8403. BR
N_Geneseq_36:V11737	+ 118.50	148.79	1.09	5711 ! Human BRCA1 DNA. Probes and p
N_Geneseq_36:V46463	+ 118.50	148.79	1.09	5711 ! Human BRCA1 om12 polymorphism
N_Geneseq_36:V46464	+ 118.50	148.79	1.09	5711 ! Human BRCA1 om12 polymorphism
N_Geneseq_36:V46449	+ 118.50	148.79	1.09	5711 ! Human BRCA1 om12 cDNA. BRCA1
N_Geneseq_36:V46458	+ 118.50	148.79	1.09	5711 ! Human BRCA1 om12 polymorphism
N_Geneseq_36:V46459	+ 118.50	148.79	1.09	5711 ! Human BRCA1 om12 polymorphism
N_Geneseq_36:V46460	+ 118.50	148.79	1.09	5711 ! Human BRCA1 om12 polymorphism
N_Geneseq_36:V60563	+ 118.50	148.79	1.09	5711 ! Tumourigenic BRCA1 allele #841

N_Geneseq_36:Q93027 + 118.50 148.79 1.09 5712 ! Gene encoding BRCA1 protein
N_Geneseq_36:T42032 + 118.50 148.79 1.09 5712 ! BRCA1 allele #77. New mutan
N_Geneseq_36:T70075 + 118.50 148.79 1.09 5712 ! Mutant BRCA1 allele #77. BR
N_Geneseq_36:T84840 + 118.50 148.79 1.09 5712 ! Human breast and ovarian ca

seq_name: N_Geneseq_36:V04079

seq_documentation_block:

ID V04079 standard; DNA; 1336 BP.

AC V04079;

DT 22-JUN-1998 (first entry)

DE Aplysia cAMP-response element binding protein 2 DNA.

KW ApCREB-2; cAMP-response element binding protein-2; snail;

KW transcription factor; memory loss; Alzheimer's disease; amnesia;

KW ischaemia; head trauma; neuronal injury; Parkinson's disease;

KW senility; therapy; ss.

OS Aplysia californica.

Key/ Location/Qualifiers

FT CDS 190..1326

FT WO9746257-A1.

PD 11-DEC-1997.

PF 03-JUN-1997; U09438

PR 03-JUN-1996; US-658811.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PI Bartsch D, Ghirardi M, Kandel ER;

DR WPI; 98-051903/05.

DR P-PSDB; W41508.

PT Enhancing long-term memory in subjects whose cAMP-responsive gene is

PT repressed - used to treat long-term memory defects, e.g. ago-related

PT memory loss, Alzheimer's disease

PS Example 2; Page 73-74; 100pp; English.

CC This polynucleotide includes a coding region for the cAMP-response

CC element binding protein-2 (see W41508), i.e. ApCREB-2, of the

CC marine snail Aplysia. ApCREB-2 is a transcription factor and

CC repressor of long-term facilitation in Aplysia neurons. It is a

CC homologue of human CREB-2 and mouse Atf-4. The polynucleotide

CC sequence was deduced from 2 independent clones isolated by a yeast

CC two-hybrid screen of an Aplysia central nervous system cDNA library.

CC The invention provides a method of enhancing long-term memory in a

CC subject whose cAMP-responsive gene expression is repressed due to

CC binding of CREB-2 to a protein or DNA associated with cAMP-responsive

CC gene expression. The method involves administering to the subject a

CC compound capable of interfering with such binding so as to derepress

CC cAMP-responsive gene expression and thereby enhance long-term memory.

CC Such compounds include anti-CREB-2 antibodies or a compound capable of

CC altering phosphorylation of CREB-2. The method is used to treat e.g.

CC age-related memory loss, Alzheimer's disease, amnesia, ischaemia,

CC shock, head trauma, neuronal injury, toxicity or degradation,

CC Parkinson's disease or senility (claimed).

CC Sequence 1336 BP; 375 A; 307 C; 328 G; 326 T;

alignment_scores:

Quality: 1923.00 Length: 378

Ratio: 5.087 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-08-656-811A-1 x V04079 ..

Align seg 1/1 to: V04079 from: 1 to: 1336

1 MetGluLeuAspLeuTrpSerGluAspPheGlnLeuAlaArgGluTrpGcl 17

|||||

190 ATGGAGCTGGACCTTTGGAGCGAAGATTTTCACTGGCCAGGAATGGG 239

|||||

17 yLeuGluMetProValValGlnThrAspGlyGlnPheGlyAspLeuLys 34

|||||

240 GCTGGAATGCCAGTCGTCAGACCGATGGCAGTTCGTGACCTCAAT 289

|||||

34 erThrSerArgHisGlyGlyAspGluSerLeuSerLeuGlnProGlnGly 50

|||||

```

290 CAACCAAGTCGTCATGGTGGCGACGAATCTTAAGTTTCAGCCCCCAGGCG 339
51 AlathrLeuLysLeuGluProPheGluGluAspValLeuGlyAlaGluTr 67
|||||
340 GCTACACTGAAGTTGGAAACCTTTGAGGAGAGATGCTCTTGGTGAGAGTG 389
67 pMetGluSerSerAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnH 84
|||||
390 GATGAGTCGTCGCGATCTCGGCTCTTTCTGGATGCTTTGGTGACAAACC 439
84 IsGluArgLeuHisProPheGluSerAsnLeuLeuGluPheThrSerLeu 100
|||||
440 ATGAGCGGCTCATCCGCTCCAGTCAAACTTTCGAGTTCACCTCTCTG 489
101 IleThrProAspAspSerThrValSerLysAspIleLeuSerSerThrLe 117
|||||
490 ATCACTCTGATGATTCGACCGGTGTCAAAGGACATCTTCACCTCAACTCT 539
117 uGlnPheProThrGlnProValAsnIleProLeuTyrAlaSerHisGlyA 134
|||||
540 TCAGTTTCCAACTCAACCACTGAACATCCCTTTATATGCAAGTCATGGGG 589
134 laGluAspPheSerAlaGluThrGluPheGluAsnHisLeuSerProPro 150
|||||
590 CCGAAGATTTCTCTGCGAGAGACTGAGTTTGAAGAACCACTTCGCCCTCA 639
151 AspSerProGluGlnValAlaProValIleAsnLeuGluProValGluLe 167
|||||
640 GATTCTCCGGAGCAGTAGCCCTGTCTAATACTAGAACCACTGTAACCT 689
167 uThrAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyGlyM 184
|||||
690 CACTGCGAGCCATATGACGGTGATCTCACTACCTGATGGCTTGTGGTGSCA 739
184 etGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsn 200
740 TCGAACTGGCTTTCAGAAAGCTTAACATTTACCGAACTAGACTTTGTGAAC 789
201 PheAsnAspSerAlaValGlySerIleGlyAlaGluGluLeuGlu 217
|||||
790 TTCATGNCAGTGTGTGGTTCAATTTGGCGGTGCTGAAGAACTTCTTGG 839
217 ySerProLeuSerValAspValGluSerThrIleSerPheSerGlyP 234
|||||
840 CTCCCCACTGTGCTGATGATGTTGGAAGTACAATATCATTTTCAGGTC 889
234 roSerSerProGluThrSerGlnSerSerIleLeuSerSerProGlu 250
890 CATCGTCCCAAGAACCCAGCCAGCAGCATCATTTGAATCAAGTCTGTAA 939
251 LeuTyrLysValIleSerThrSerSerIleAspAlaSerLysArgPheSe 267
940 TTGTACAAAGTTATCTTACCTCGTCCATTCATGCACTAAGCGTTTCTC 989
267 rProTyrSerArgSerSerLysSerLysGlnSerValLysThrSerAsp 284
990 TCCATACTCTGTTCTCCCAAGTCCAAGCAATCTGTCAAGACTTTCAGAGC 1039
284 laLysAlaProArgLysThrArgThrProAlaGlnProValProGluHis 300
1040 CTAAGGCACCTCGTAAACAGGAGACACCGCGCGACCTGTGCCAGAACAT 1089
301 ValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsnLy 317
1090 GTCATCATGGAAACATTTGGACAAAAGGACAGAAAGAGCTTCAGAACAA 1139
317 sAsnAlaIleArgTyrArgMetLysLysLysGlyGluAlaGlnGlyI 334
1140 GAATGCTGCCATTAGGTATAGTATAGTGAAGAAGAGGGAGGCTCAGGGCA 1189
334 leLysGlyGluGluGlnGluLeuGluLeuAsnThrLysLeuLysThr 350
1190 TCAAGGGGAGGAAACAGAAATTAGAAGAACTCAACACAAAGCTTTAAGACT 1239

```

```

351 LysValAspAspLeuGlnArgGluIleLysTyrMetLysAsnLeuMetG1 367
|||||
1240 AAGTCTGATGACTTGCAAGAGAAATCAAGTACATGAAAAATTTAATGGA 1289
367 uAspValCysLysAlaLysGlyIleGlnLeuLys 378
|||||
1290 AGATGTTTGCAGGCGGAAAGCTATTTCAGCTTAAA 1323
seq_name: N_Geneseq_36:T77784

```

seq_documentation_block:

```

ID T77784 standard; cDNA; 669 BP.
AC T77784;
DT 01-OCT-1997 (first entry)
DE cDNA clone 59 encoding NIP-1.
KW NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;
KW cell division; proliferation; antibody; Ab; detection;
KW malignant cell growth; ss.
OS Homo sapiens.
PN W09640917-A1.
PD 19-DEC-1996.
PF 07-JUN-1996; U09504.
PR 07-JUN-1995; US-478408.
PA (UYA ) UNIV YALE.
PI McPherson SMG, Snyder MP;
DR WPI: 97-077270/07.
DR P-PSDB; W21733.
PT New nucleic acid encoding nuclear mitotic appts. interacting
PT proteins - useful for modulating cell division and proliferation and
PT in diagnosis.
PS Claim 28; Page 57-58; 78pp; English.
CC The sequences given in T77784-86 encode NuMA binding proteins (NuMA -
CC nuclear mitotic apparatus). These protein sequences were identified
CC using the fusion proteins given in W21731-32. Compounds which interfere
CC with the interaction of NuMA with a known NIP (NuMA interacting protein)
CC are used to modulate cell division and/or proliferation. Ab, raised
CC conventionally using NIP-1 or -2 as immunogen, are used to detect NIP
CC (or their complexes) and to block their activity for diagnostic or
CC therapeutic use, e.g. to detect defective NuMA or NIP which may be
CC markers for aberrant (including malignant) cell growth (which can also
CC be detected by nucleic acid sequencing). Also where malignancy is
CC related to defects in NuMA or NIP, it can be treated by administration
CC of the appropriate functional protein. The protein encoded by this
CC sequence is rich in proline residues.
SQ Sequence 669 BP; 131 A; 232 C; 191 G; 115 T;

```

alignment_scores:

```

Quality: 137.50 Length: 148
Ratio: 1.599 Gaps: 5
Percent Similarity: 58.108 Percent Identity: 29.730

```

alignment_block:

```

US-08-656-811a-1 x T77784 ..

```

```

Align seg 1/1 to: T77784 from: 1 to: 669

```

```

234 ProSerSerProGluThrSerGlnSerSerIleLeuGluSerSerProG1 250
|||||
28 CCCCTCCCTCCCTGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTG 77
250 uLeuTyrLysValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
|||
78 CTGG.....GATCTCTGGACTGTCGTGG 100
267 erProTyrSerArgSerSerLysSerLysGlnSerValLysThrSerAsp 283
:: |||
101 CCATCTACTGCGCAACGAGCGCGGAGGAGGAGTGGGATGCCGCT 150
284 AlalysAlaProArgLysThrArgThrProAlaGlnProValProGlu... 299
|||
151 CTGCCCCCGGCACAGCAGCCCCCTCCCTCTTCTCCACCTCAACCTTCTCG 200

```



```

300 .....HisValIleMetGluHisLeuAspLysLysAspA 311
      |||      :|: |||:|||||:
201 COTGCCCCCTACCCACATCTCCACACCGAGGGGACCGCAAGCAAA 250
      |||:|||||:
311 tGLysLeuGlnAsnLysAsnAlaIleArgTyrArgMetLysLys 327
      :|||:|||||:|||||:|||||:|||||:|||||:
251 AGAAGAGAGACAGAACAGTCGGCGGCTCTGAGGTACCGCCGCGAAGC 300
      :|||:|||||:|||||:|||||:|||||:|||||:
328 LysGLyGluAlaGlnGlyIleLysGLyGluGlnGluGluL 344
      :|||:|||||:|||||:|||||:|||||:|||||:
301 GAGAGGGTGAGCCCTGGA.....GCGAGTGCCAGGGGCTGGAGGCAC 344
      :|||:|||||:
344 euAsnThrLysLeuLysThrLysVal.AsaspLeuGlnArgGluIleLys 360
      ||| :|||:|||||:|||||:|||||:|||||:|||||:
345 GGAATCGCGAGCTGAAGAACGGGCAAGAGTCCGTGGAGCGCGAGATCCA 394
      :|||:|||||:|||||:|||||:|||||:|||||:
360 stYrMetLysAsnLeuMetGluAspValCysLysAlaLys 373
      :|||:|||||:|||||:|||||:|||||:|||||:
395 GTACGTCAAGGACCTGCTCATCGAGGTTTACAGGCCCGG 434

```

```
seq_name: N_Geneseq_36.V29204
```

```
seq_documentation_block:
```

```
ID V29204 standard; cDNA; 3190 BP.
```

```
AC V29204;
```

```
DT 10-AUG-1998 (first entry)
```

```
DE Gene encoding the mouse cAMP regulatory element binding protein.
```

```
KW Mouse cAMP regulatory element binding; mCREBa; antibody; inhibition;
```

```
KW mCREBa modulator; ss.
```

```
OS Mus sp.
```

```
FT Key Location/Qualifiers
```

```
FT CDS 304..1869
```

```
FT /tag= a
```

```
FT /product= "mCREBa"
```

```
FN WO9812910-A2.
```

```
PD 02-APR-1998.
```

```
PF 26-SEP-1997; U17288.
```

```
PR 27-SEP-1996; US721684.
```

```
PA (ICOS-) ICOS CORP.
```

```
PI Keegan KS;
```

```
DR WPI: 98-230326/20.
```

```
DR P-PSDB; W37934.
```

PT Murine cAMP regulatory element binding protein A, mCREBa - useful for systematic analysis of structure and function of mCREBa and identification of those molecules with which it will react

PS Claim 3; pages 26-29; 41pp; English.

CC This is the nucleotide sequence of the gene encoding the mouse cAMP regulatory element binding (mCREBa) protein. The products of the mCREBa such as antibodies can be used for the recombinant production of the protein, to identify novel genes encoding binding partner polypeptides for mCREBa, for the systematic analysis of the structure and function of mCREBa, and identification of those molecules with which it will react, to identify inhibitors of mCREBa binding to other natural binding partners and to generate rodents that fail to express a functional mCREBa or express a variant mCREBa, useful as models for studying the activities of mCREBa and mCREBa modulators in vivo.

```
SQ Sequence 3190 BP; 729 A; 939 C; 778 G; 744 T;
```

```
alignment_scores:
```

```
Quality: 132.00 Length: 423
```

```
Ratio: 0.635 Gaps: 22
```

```
Percent Similarity: 49.173 Percent Identity: 21.513
```

```
alignment_block:
```

```
US-08-656-811A-1 x V29204 ..
```

```
Align seg 1/1 to: V29204 from: 1 to: 3190
```

```
19 GluMetProValGlnThrAspGlyGlnPheGlyAspLeuLysSerTh 35
```

```
|||||: |||:||||| ||| :|||:|||||:|||||:|||||:|||||
```

```

325 GACGACGAGCGTCTCGACAGTGGGACCCCAAGCTGAGCGAGCTG..... 366
35 rSerArgHisGlyGlyAspGluSerLeuGlnProGlnGlyValat 52
   |||:|||||:
367 .TCAGAGCCCGGAGAGACTGAGCCCTCATGTACACACAGCCTTCTCGG 415
52 hrLeuLysLeuGluProPheGluGluAspValLeuGlyAlaGluTyrMet 68
   ||| |||:|||||:|||||:|||||:|||||:
416 AGCTC...CTAGACGAGTTCCTCCAGAGCTCTCTGGGT..... 450
69 GluSerSerAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisG1 85
   :|||:|||||:
451 .....CAGCTCTCGAGTGAC..... 465
85 uArgLeuHisProPheGluSerAsnLeuGluPheThrSerLeuIlet 102
   |||:|||||:
466 .....CCTTCTCTCTCA..... 477
102 hrProAspAspSerThrValSerLysAspIleLeuSerSerThrLeuGln 118
   :|||:|||||:
478 .....GAGAAGAGCGAGTCAATGGAGGTGGAGCCATCT..... 510
119 PheProThrGlnProValAsnIleProLeuTyrAlaSerHisGlyAlaG1 135
   |||:|||||:|||||:
511 ...CCAACATCACCGCG.....CCTCTCATCCAGGCTGAACACAGCTA 551
135 uAspPheSerAlaGluThrGluPheGluAsn.....HisLeuSerP 149
   :|||:|||||:
552 CTCCTGTGAGCGAGGAGCCCGGACCTCAGTCACCATTTACCCATCGCGCTA 601
149 roProAspSer.....ProGluGlnVal..... 156
   |||:|||||:
602 CCAGCGACAGCTTCAATGACGAGGAGGTGGAGAGTGAAAAAATGGTACCTG 651
157 .....AlaProValIleAsnLeuGluProValGluLeuTh 168
652 TCTACAGAGTTTCTCTTACCTACCTACCATCAAGAAGACCAATC..... 693
168 rAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyGlyMetG 185
   ||| :|||:|||||:
694 .....ACAGAGGAGCAGGCCCGCGGACTT..... 717
185 luLeuAlaSerGluSerLeuThrGluLeuAspPheValAsnPhe 201
   :||| :|||:|||||:|||||:
718 ..GTCCCTTCTGTCTGCTGACATCAGCATCACAGCCATTTCC...ACTCCTTT 762
202 Asn.....AspSerAlaValGlySerIleGlyAl 212
   :|||:|||||:
763 GAAAAGAAAGAGTCCCTCTGGATATGAATGTGGGGGGGACTCCTCATG 812
212 aGluGluLeuLeuGlySer...ProLeuSerValAspAspValGluSert 228
   :|||:|||||:
813 CCAGAGCGCTTATTCCTAAGATTAAAGCTGGAGCCCAAGAGTGATCAGT 862
228 hrIleSerPheSerGlyProSerSerProGluThrSerGlnSerIle 244
   :|||:|||||:|||||:
863 TCTTAAACTTCTCC...CGAAAAGAGCCTCGGTGGATCAACTGCACCTTA 909
245 IleGluSerSerProGluLeuTyrLysValIleSerThrSerSerIleAs 261
   :|||:|||||:
910 CCACCAACACACCCACAGTAGTCACAGCAGTACTCTGAGGGGAGCTTGAG 959
261 pAlaSerLysArgPheSerProTyrSerArgSerSer..... 273
   :|||:|||||:
960 CCCCACACCCAGCGCTGCATCCCTTCAGCCTGTCTCAGGCCCCACAGCCCTG 1009
274 .....LysSerLysGlnSerValLysThrSerAsp...Ala 284
1010 TCAGAGCCCATGCCCGGGGCCCTCTGCTTGTCCACATCTCCTCTCCTC 1059
285 LysAlaProArgLysThrArg..... 291
   |||:|||||:
1060 ACAGTCCACATAAAGCTGACGGGATCGGCCCCCTCGTCTCTGACAGAAGA 1109

```

292ThrProAlaGlnProValProGluHisValI 302
 1110 GGAGAGAGGAGCCCTGCTGCGGAGGCTATCCCATCCACCAAGCTGC 1159
 302 leMet.....GluHisLeuAspLysLysAspArgLysLysLeu 314
 1160 CTCTGACAAAATCTGAGGAGAGGCCCTGAAGAAAATCCGGAGAAAGATC 1209
 315 GluAsnLysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGluAl 331
 1210 AAGAAATAGATTTCTGCCCAAGAAGCAGGAGAAAGAAAGAAAGAAATACAT 1259
 331 aglncllyleLysGlyGluGluGlnGluLeuGluGluLeuAsnThrLysL 348
 1260 GGACAGCCTGGAGAAAAAGTGGAGTCTGTTCACACTCAGAACTTGGAGC 1309
 348 euLysThrLysValAspAspLeuGln.....MetGluAspValCy 370
 1310 TTCGGAAGAAGTGGAGGTGCTGGAGAACACCAATAGACTCTCCTTCAG 1359
 358 GluIleLysTyrMetLysAsnLeu.....MetGluAspValCy 370
 1360 CAACCTTCAGAGCTTCAGACTTTGGTGATGGGGAAGGTCTCGAAGCTG 1409
 370 sLysAlaLysGlyIleGln 376
 1410 CAAGTTACCTGGCACACAG 1428

seq_name: N_Geneseq_36:T14025
 seq_documentation_block:
 ID T14025 standard; cDNA; 798 BP.
 AC T14025;
 DT 09-JUL-1996 (first entry)
 DE Drosophila dCREB1 cDNA.
 KW Enhancer-specific activator; dCREB1; long-term memory;
 KW transgenic animal; insect; ds.
 OS Drosophila sp.
 PN W09611270-Al.
 PD 18-APR-1996.
 PF 06-OCT-1995; U13198.
 PR 07-OCT-1994; US-319866.
 PR 21-DEC-1994; US-361063.
 PA (COLD-) COLD SPRING HARBOR LAB.
 PI Regulski M, Tully TP, Yin JC;
 DR WPI; 96-209851/21.
 DR P-PSDB; J91295.
 PT Novel cyclic 3',5'-adenosine monophosphate responsive
 PT transcriptional activator gene - used in stimulation and enhancement
 PT of longterm memory
 PS Claim 31: Page 100-101; 160pp; English.
 CC The Drosophila CREB1 gene (T14025) codes for an enhancer-specific
 CC activator (R91295) that appears to be a member of the CREB/ATF
 CC family and that is associated with long-term memory. The gene was
 CC isolated by a DNA-binding expression screen of a Drosophila head cDNA
 CC library using a probe contg. 3 CAMP-responsive element sites. dCREB1
 CC and dCREB2 (see also T14024) genes can be used to manipulate memory,
 CC and transgenic insects carrying the genes can be used to assess the
 CC effects of drugs on long-term memory formation.
 SQ Sequence 798 BP; 245 A; 173 C; 175 G; 205 T;

alignment_scores:

Quality: 128.50 Length: 310
 Ratio: 0.824 Gaps: 18
 Percent Similarity: 50.323 Percent Identity: 24.516

alignment_block:

US-08-656-811A-1 x T14025

Align seg 1/1 to: T14025 from: 1 to: 798

91 GluSerAsnLeuLeuGluPheThrSerLeuIleThrProAspAspSerTh 107
 13 GAAAATATGTTTCTACTTTTCACATCGTTA.....GATGCTGCTAC 53
 107 rValSerLys..... 110
 54 CGGTACCAACCAACACCGGTGAATTTCTTAATGAATGAATCTCCAAGGCAAG 103
 111AspIleLeuSerSerThrLeuGlnPheProThrGlnProVal 124
 104 AAGCCGGTGACTTAATGTTGAGTACTCTGGATTTCAACATTTATGGCGAA 153
 125 AsnIleProLeuTyrAlaSerHisGlyAlaGluAspPheSerAlaGluTh 141
 154 AACCTG.....GCAGATGATTTCACAGACCTCG... 180
 141 rGluPheGluAsnHisLeuSerProAspSerProGluGlnValAlap 158
 181GCTTCACAGCTTCGGAGGACAAAGATGACTC 211
 158 ro...ValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
 212 CTTTCGTGTGTATACCACTGTTTTTGAATCCGCTTCAAGAAC..... 255
 174 ValIleSerProAspGlyLeuLeuGlyMetGluLeuAlaSerGluSe 190
 256ACCGAAGATACCTTCTAGGA.....GATATCGACAATGTTGG 293
 190 rLeuThrPheThrGluLeu...AspPheValAsnPheAsnAspSerAlav 206
 294 TATTGTTGACACGGAGTGAAGGAGATGTTGCGATTGTTGACTCGGAAA 343
 206 alGlySerIleGlyGlyAlaGluLeuLeuGlySerProLeu..... 220
 344 TCAATAAC.....GGCACTCTCATCAACGAG 369
 221 ...SerValAspValGluSerThrIleSerPheSerGlyProSerSe 236
 370 GAACAAAAGGATGATTGGAATTACT.....TC 398
 236 rProGluThrSerGlnSerSerIleIleGluSerSerProGluLeuTyrL 253
 399 AAGATCCCAAGTCCACCTCAGCTCTCTTGTGCTCG.....A 433
 253 ysValIleSerThrSerSerIleAspAlaSerLysArg...PheSerPro 268
 434 AATCGACTTCTGCTTCTCCAGCTGATGCTGCCGCTGCATGTCGCAAGTCCT 483
 269 TyrSerArgSerSerLys.....SerLysGlnSerValLysThrSe 282
 484 TCGTCATCGTCTTGTAAAGATCCTATTCTTCTGCTCAGCTAGAAACTAC 533
 282 rAspAlaLysAlaProArgLysThrArg.....ThrPro..... 293
 534 GGGTTCGGATGCTCCAAGAAAGATAAGCTGGGCTGCACCCCTTACACTA 583
 294AlaGlnProValProGluHisValIleMetGluHisLeu 306
 584 GAAACACAGAGAAACATCCATTACTCG...GTCATTCCAAGGGTCAG 630
 307 AspLysLysAspArgLysLysLeuGlnAsnLysAsnAlaIleArgTy 323
 631 GATGTTCTTCTATGAAAGGGCAAGAACACTCAGGCCGCAAGATC 680
 323 rArgMetLysLysGlyGluAlaGlnGlyIleLysGlyGluGlnG 340
 681 AAGAGCCAGAAAAATGGAAGAATG.....TCCC 709
 340 luLeuGluGluLeuAsnThrLysLysValAspAspLeuGln 356
 710 AACTTGAGAAAAAGTGTCAAAGCTTGTGAAGGAAACGACGACTTGAAA 759
 357 ArgGluIleLysTyrMetLysAsnLeuMet 366


```

2730 AGTCCTTGATAGAGCGAGCCCTGTTCACTTATACACTGGATGAAAAG 2779
306 LeuAspLysLysAspArgLysLysLeuGlnAsnLysAsnAlaAlaIleAr 322
2780 GTGACGAGTGTGGCGGCTTCTCTC.....CTCAA 2811
322 gTyrArgMetLysLys 327
2812 ATATCAAGTGAAGCAG 2827
seq_name: N_Geneseq_36.V69720

```

seq_documentation_block:

```

ID V69720 standard; cDNA; 4225 BP.
AC V69720;
DT 01-MAR-1999 (first entry)
DE Tumour rejection antigen precursor MAGE-C1 cDNA.
KW MAGE-C1; human; tumour rejection antigen precursor; TRAP;
KW therapy; diagnosis; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 257..3685 /*tag= a
FT WO9849184-A1.
FN 05-NOV-1998.
PD 24-APR-1998; U08493.
PR 25-APR-1997; US-845528.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Boon-Faleur T, De Smet C, Lucas S;
DR MPI; 99-024041/02.
DR P-PSDB; W81546.
DT Tumour rejection antigen precursors - used for determining presence
PT of cytolytic T cells specific for complexes of a human leukocyte
PT antigen
PS Claim 1; Page 48-50; 84pp; English.
CC This nucleotide sequence encodes novel human tumour rejection
CC antigen precursor (TRAP) MAGE-C1 (see W81546). MAGE-C1 is a novel
CC member of the MAGE family that may be recognised by cytotoxic T
CC cells, leading to lysis of the tumour cells which express it.
CC MAGE-C1 and MAGE-C2 (see W81547) are expressed in a variety of
CC tumours and in normal testis cells, but not by other normal cells.
CC The MAGE-C1 cDNA was isolated from a melanoma LB373-MEL cDNA by
CC PCR amplification (see V69732-33). It shows homology to MAGE-A1
CC cDNA (see V69719). The MAGE-C1 gene was localised to Xq26-q27.
CC MAGE-C1 and MAGE-C2 cDNAs (see V69726) are claimed, as are:
CC expression vectors; transformed or transfected cell lines (e.g. COS
CC and CHO); an isolated TRAP encoded by the cDNAs; a kit useful in a
CC PCR based assay; a method for determining expression of a MAGE-C1
CC gene using the kit; a polypeptide comprising a number of tumour
CC rejection antigens derived from MAGE-C1 or MAGE-C2; and a polypeptide
CC comprising at least one tumour rejection antigen derived from
CC MAGE-C1 or MAGE-C2 and at least one other tumour rejection antigen.
CC MAGE-C1 and MAGE-C2 can be used in a method for determining the
CC presence of cytolytic T cells specific for complexes of a human
CC leukocyte antigen (HLA).
SQ Sequence 4225 BP; 871 A; 1198 C; 923 G; 1233 T;

```

alignment_scores:

```

Quality: 124.50 Length: 372
Ratio: 0.688 Gaps: 17
Percent Similarity: 48.656 Percent Identity: 22.043

```

alignment_block:

```

US-08-656-811a-1 x V69720 ..
Align seg 1/1 to: V69720 from: 1 to: 4225
5 LeuTrpSerGluAspPheGlnLeuAlaArgGluTrpGlyLeuGluMetPr 21
|||||.....|
2102 CTTAGGGGGAGGAGTTCAGTCTTCT.....CTCCAGAGGCC 2139

```

```

21 oValValGlnThrAspGlyGlnPheGlyAspLeuLysSerThrSerArgH 38
|||||.....|
2140 TGTGAGCATCTGCTCC.....TCCTCCACTCCATCCAGTCTTCCCC 2180
38 isGlyGlyAspGlySerLeuSerLeuGlnProGlnGlyAlaThrLeuLys 54
|||||.....|
2181 AGAGTTTCCCTGAGAGATTCTCAGAGTCTCTCTGAGGGGCTGTC..... 2224
55 LeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerSe 71
|||||.....|
2225 CAGTCTCTCTCCATAGTCTCTCAGAGCCCTCTGAGGGGATGCACTCCA 2274
71 rAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeuH 88
|||||.....|
2275 ATCTCTCTCTCCAGAGTCTCTGAGAGTGTCTCTGAGGGGAGGATTCCTGT 2324
88 isPropheGluSerAsnLeuLeuGluPheThrSerLeuIleThrProasp 104
|||||.....|
2325 CTCCTCTCCAA.....ATTCTCTCAG 2344
105 AspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPhePro.. 120
|||||.....|
2345 AGTCTCTCTGAGGGAGAGACTCCCTGTCTCT...CTCCATTTTCCCTCA 2391
121 .....ThrGlnProValAsnIleProLeuT 129
|||||.....|
2392 GAGTCTCTCTGAGTGGGAGGACTCCCTCTCTCTCTCCACTTCTCTCAGT 2441
129 yrAlaSerHisGlyAlaGluasphe..... 137
|||||.....|
2442 TTCTCTCTCAGGGG...GAGGACTTCCAGTCTCTCTCCAGAGTCTCTGTG 2488
138 .....SerAlaGluThrGluPhe.....G1 144
|||||.....|
2489 AGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2538
144 uAsnHisLeuSerProProAspSerProGluGlnValAlaProValIleA 161
|||||.....|
2539 GAGTCTCTCAGAGTCTCTCTGAGGGGCTCTCTCTCTCTCTCTCTCTCT 2581
161 snLeuGluProValGlu.....Leu 167
|||||.....|
2582 ..CAGAGACCTGTCTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2629
168 ThrAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyMe 184
|||||.....|
2630 CAAGTTCCCATGAGAGTCTCTCAGAGTCTCTCTCTCTCTCTCTCTCT 2665
184 tGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsnP 201
|||||.....|
2665 ..... 2665
201 heAsnAspSerAlaValGlySerIleGlyAlaGluGluLeuLeuGly 217
|||||.....|
2666 .....GGGCTCTGCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2689
218 SerProLeuSer.....ValAspAspValGluSerThrIleSerPheSe 232
|||||.....|
2690 AGTCTGTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2739
232 rGlyPro.....SerSerProGluThrSerGlnSerSerIleIleGluS 247
|||||.....|
2740 TTCCTCTGTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2789
247 erSerProGluLeuTyrLysValIleSerThrSerSerIleAspAlaSer 263
|||||.....|
2790 GTTCCCTGAGAGTCTCTCTCCAGAGTCTCTGTGATCTCTCTCTCTCTCT 2839
264 LysArgPheSerProTyrSerArgSerSer..... 273
|||||.....|
2840 ACTTCATTGAGCCCATTCAGTGAAGAGTCCAGAGCCGCCAGTAGATGAATA 2889
274 ....LysSerLysGlnSerValLysThrSerAspAlaLysAlaProArgL 289

```

```
2890 TACAAGTTCCTCAGACACCTTCCTAGAGGTGATTCCTTGACGACGCG 2939
      :::::::::::::::::::: |||||::: :::
289 yThrArgThrProAlaGlnProValGluHisValIleMetGluHis 305
      ::::: ::::::::::::::: ::::::::::: |||
2940 AGTCCTTGATAGAGCGAGCGCCTGTTGTCATATACATGGATGAAAG 2989
      ::::::::::::::: |||
306 LeuAspLysLysAspArgLysLysLeuGlnAsnLysAsnAlaIleAla 322
      ::::::::::::::: |||
2990 GTGGACGAGTGGCGGCTTCTTC.....CTCAA 3021
      ::::::::::::::: |||

322 gTyrArgMetLysLys 327
      ::::::::::::::: |||
3022 ATATCAAGTGAAGCAG 3037
      ::::::::::::::: |||

seq_name: N_Geneseq_36:T17502
seq_documentation_block:
ID T17502 standard; cDNA; 5914 BP.
AC T17502:
DT 02-OCT-1996 (first entry)
DE Mutated BRCA1 coding sequence from PM20.
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 120..5711
FT /*tag= a
FT /product= BRCA1 mutant
FT mutation 3238
FT /*tag= b
FT /note= "G to A mutation"

WO9605306-A2.
PD 22-FEB-1996. U10202.
PF 11-AUG-1995; US-289221.
PR 12-AUG-1994; US-300266.
PR 02-SEP-1994; US-308104.
PR 16-SEP-1994; US-348824.
PR 29-NOV-1994; US-409305.
PR 24-MAR-1995; US-483553.
PR 07-JUN-1995; US-480784.
PR 07-JUN-1995; US-480784.
PA (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
DR WPI: 96-139702/14.
DR P-PSDB: R81540.
DT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
PS Claim 1: 218pp; English.
CC T17439-T17453 and T17455-T17529 represent mutations of the human breast
CC and ovarian cancer predisposing gene (BRCA1) (for wild type see T17438).
CC Proteins encoded by these mutations (see R81493-R81497 and R81499-R81546)
CC can be used as immunogens for antibody production. These mutant BRCA1
CC genes have at least 1 mutation or polymorphism in comparison to the wild
CC type sequence. By detecting a germline alteration in the wild type BRCA1
CC gene, a predisposition for breast and ovarian cancer can be diagnosed.
CC In one method, BRCA1 mRNA isolated from a tissue sample from a subject
CC has a probe, corresponding to a fragment of the wild type sequence (or an
CC allele-specific probe for one of these mutations), added to it. The
CC conditions allow for hybridisation of the probe to the mRNA, and any
CC hybridisation which occurs is detected. Alternatively the BRCA1 gene in
CC the tissue sample is isolated, and a shift in electrophoretic mobility of
CC single stranded DNA from the sample on a non-denaturing polyacrylamide
CC gel indicates a mutation. These methods of detection can also diagnose a
CC lesion neoplasia associated with the BRCA1 locus. The methods may be
CC used in gene therapy, protein replacement therapy and protein mimetics,
CC and may be used to screen for drugs in cancer therapy.
CC Sequence 5914 BP; 2007 A; 1156 C; 1315 G; 1436 T;
```

```
alignment_scores:
  Quality: 121.50      Length: 436
  Ratio: 0.620        Gaps: 19
  Percent Similarity: 44.954      Percent Identity: 19.037

alignment_block:
US-08-656-811A-1 x T17502 ..
Align seg 1/1 to: T17502 from: 1 to: 5914

13 AlaArgGluTrpGlyLeuGluMetProValValGlnThrAspGlyGlnPh 29
   ::::::::::: ::::::::::: |||||::: ::: ||::::: ||:::
2637 AGTCGGGAACAAGCATAGAAATGGGAAGAAAGTGAACCTTGATGCTCAGTA 2686
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
29 eGlyAspLeuLysSerThr..... 35
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2687 T.....TTGCAGATACATTCAAGGTTTCAAGGCCCGCAGTCATTGTGCTC 2730
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
36 .....SerArg 37
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2731 CGTTTTCAATCCAGGAAATGCAGAGAGGAATGTGCAACATTTCTCTGCC 2780
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
38 HisGlyGlyAspGluSerLeuSerLeuGlnProGlnGlyAlaThrLeuLys 54
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2781 CACTCTGGG.....TCCTTAAGAAACAAGTCCAAAGAGTCACTTTTGA 2824
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2825 ATGTGAACAAGGAAGAAAT..... 2846.
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
71 eAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2847 .....CAAGGAAAGAATGAGTCTAATATC 2870
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
88 HisProPheGluSerAsnLeuLeuGluPheThrSerLeuIleThrProAs 104
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2871 AAGCCTGTACAGACA..... 2885
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProt 121
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2886 .....GTTAATATCACTGCAGGCTTTCCCTG 2910
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
121 hr.....GlnProValAsnIleProLeuTyrAlaSerHis 132
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2911 TGGTTGGTCAGAAAGATAAGCCAGTTGATAATGCCAAATGTAGTATCAAA 2960
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnH1 146
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
2961 GGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAGAGGCAACGAAACTGG 3010
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
146 sLeuSerProProAsp.....SerProGluGlnValA 157
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
3011 ACTCATTACTCCAAATAACATGGAGCTTTTACAAAACCCATATCGTATAC 3060
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
157 lAsProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
3061 CACCACCTTTTCCCATCAAGTCATTGTTTAAACATAATGTAAGAAAAAT 3110
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
174 ValIle.....SerProAspGlyLeuLe 191
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
3111 CTGCTAGAGGAAACACTTTGAGGNACATTCATGTCTCACCTGAAGAGAAAT 3160
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThrPheThr..... 194
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
3161 GGGAAATGAGACATTCACCAAGTACACTGAGCACATATTAGCCGTAATAACA 3210
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
195 .....GluLeuAspPheValAsnPheAsnAspSerAlaValGlySerIle 209
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
3211 TTAGACAAAAATGTTTTTAAAGAAGCAACTCAAGCAATATTATTAAAGTA 3260
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
210 GlyGlyAlaGluLeuLeuGlySerProLeu.....Se 221
   ::::::::::: |||||::: |||||::: |||||::: |||||::: |||||:::
```



```

626 .....AGCGCGGCACCACTTGCCTCCCAACAGATCCC 556
273 erLysSerLysGlnSerValLysThrSerAspAlaLysAlaProArgLys 289
657 GGTGCAGC.....ACCGCGGCTGCAAGCCCTCAAGGAAG 691

290 ThrArgThrProAla.....GlnProValProG 299
692 AGCGCGAGACCTGCGGAGATGCGCGGAGACGCCGCCCTGTCCCT 741
299 LuHisValIleMetGluHisLeuAsp.LysLysAspArgLysLysLeuG1 315
742 ATCGACATGAGTCTCAGGAGCGGATCAAGGAGAGAGAGAGCGGATGAG 791
315 nAsnLysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGluAlaG 332
792 GAACCGCATTCGCCCTCCAAAGTCCGGAAGGAGGAGCTGGAGCGATCG 841
332 InGlyIleLysGlyGluGluGlnGluLeuGluGluLeuAsnThrLysLeu 348
842 CTCGGCTAGAGAAAGTGAACCTTGAAGCGCAAACTCCGAGCTG 891
349 LysThrLysValAspLeuGlnArgGluIleLysTyrMetLysAsnLe 365
892 GCATCCAGCGCAACATGCTCAGGAACAGGTGGCACAGCTTAAG..... 936
365 uMetGluAspValCysLysAla.....LysGlyIleGlnLeuLysMet 379
937 ....CAGAAAGTCATGAGGCATCCACGAGTGGTGGTCCCAACTCATGCTA 981

seq_name: N_Geneseq_36:Q13674

```

```

seq_documentation_block:
AC Q13674 standard; DNA; 3518 BP.
DE 28-NOV-1991 (first entry)
DE E2A.E12 transcript.
KW Immunoglobulin; enhancer-binding factor; chromosomal translocation;
KW leukemia; fusion protein; ss.
OS Homo sapiens.
FH Key . Location/Qualifiers
FT cds 31..1992 /*tag= a
FT FT /label= E2Aalpha
FT FT /note= "mol. wt. 67.6 kD"
FT poly_a_signal 2513..2518 /*tag= b
FT FT
PN W09113172-A.
PD 05-SEP-1991.
PF 22-FEB-1991; U01168.
PR 23-FEB-1990; US-484063.
PA (STRD ) LELAND STANFORD JR UNIV.
PA (WHIT-) WHITEHEAD INST BIOMED RE.
PI Cleary ML, Wellentin JD, Baltimore D, Murre C, Mccaw P;
DR WPI; 91-281484/38.
DR P-PSDB; R13950.
PT Detection of t(1;19) break-point-associated genes E2A and p11 -
PT in chromosomal translocation, and prods, useful in diagnosis and
PT therapy of human neoplasm, esp. acute lymphoblastic leukaemia
PS Disclosure; Fig 4A; 104pp; English.
CC The cDNA contains a single ORF encoding 654 amino acids and 2.4 kb
CC of 3' untranslated mRNA. The 3' untranslated region is not fully
CC represented due to discontinuous reproduction in the specification.
CC With the exception of a single in-frame insertion of 3 bases,
CC nucleotides 678-2015 are identical in sequence to the nucleotides
CC 9-1343 of the E12 isolate of the E12 gene (Q13670). The inserted
CC triplet results in the addition of glutamine 390.
CC See also Q13669-75.
SQ Sequence 3518 BP; 703 A; 1166 C; 1020 G; 629 T;

```

alignment_scores:

```

Quality: 120.50 Length: 411
Ratio: 0.600 Gaps: 13
Percent Similarity: 48.905 Percent Identity: 20.195

alignment_block:
US-08-656-811A-1 x Q13674 ..

Align seg 1/1 to: Q13674 from: 1 to: 3518

4 AspLeuTrpSerGluAspPheGln.....LeuAlaArgG1 15
706 GAGCTCTGAGTCCCCCGGCGGCGGCTTCGGGCCCATCTGGGTGG 755
15 utrGlyLeuGluMetProValValGlnThrAspGlyGlnPheGlyAspL 32
756 GGGCTCATCTCCGCTCCCTCCGCGGCTAGCGGCGGCTGGCGAGCA 805
32 euLysSerThrSerArgHisGlyGly.....AspGluSerLeuSer 45
806 GTGGAAAGCAGCAGCAGCTTTGGTGGCTGCACACGACGAGCGTATGGGC 855
46 LeuGlnProGlnGlyAlaThrLeuLys..... 54
856 TACCAGCTCCATGGACGACAGGTGAACGGTGGCTCCCATCTCCATCCTC 905
55 .....LeuGluProPheGluG 60
906 CTTCTCCTCAGCCCCCGGAGCCACGCTACGGCGGCTCTCCAGCCACACGC 955
60 luAspValLeuGlyAlaGluTrpMet.....GluSerSerAspLeu 73
956 CGCCTGTCAGCGGGCGGACAGCCTCTCGGCTCCGAGGAGCACACAGCT 1005
74 GlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeuHisProPh 90
1006 GCGAGCTCCGGGATGCCCTCGGCAAGCACCTGCGCTCATCTACTACCCC 1055
90 eGluSerAsnLeuLeuGluPheThrSerLeuIleThrProAspAspSerT 107
1056 GGATCACTCAAGCAATAACTTCTCGTCC.....AGCCCTTCTACCCCG 1099
107 hrValSerLysAspIleLeuSerSerThrLeuGlnPheProThrGlnPro 123
1100 TGGGCTCCCCCAGGCGCTGGCAGGAACGTCACAGTGGCTCGAGCAGGA 1149
124 ValAsnIleProLeuTyrAlaSerHisGlyAlaGluAspPheSerAlaG1 140
1150 GCGCCCGGTGCTTATCGCCCGACGACGCGGGGTCTCCACGGCGCTGCA 1199
140 uThrGluPheGluAsnHisLeuSerProProAspSerProGluGlnValA 157
1200 GAGTAAGATAGAAGACCACTG.....GACGAGGCCA 1231
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
1232 TCCAGCTGTCTCCGACGCCACGCGCTGGGC...ACAGCGGCGGACATGCA 1278
174 ValIleSerPro.....AspGlyLeuLeuGlyGlyMe 184
1279 AGCTGTCTCTGGCCACCGGCGCTGGCTCAGGTTTCACCGGCCCAT 1328
184 tGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsnP 201
1329 GTCGTGGTGGCGGCAC..... 1347
201 heAsnAspSerAlaValGlySerIleGlyGlyAla.....GluGluLeu 215
1348 .....GCAGCGCTGGTGGAGGACCAACCCCGGAGCGGC 1383
216 LeuGlySerProLeuSerValAspValGluSerThrIleSerPheSe 232
1384 CTCGAGGCGGACCAACGCTCATGTGCAACACGCGGCGGCTC..... 1425

```



```

2731 CGTTTCAATCCAGGAATGCAGNAGAGGAATGTGCAACATTTCTCTGCC 2780
38 HisGlyGlyAspGluSerLeuSerLeuGlnProGlnGlyAlaThrLeuLeu 54
|||||
2781 CACTCTGGG.....TCCTTAAGAAACAAAGTCCAAAAGTCACCTTTTGA 2824
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
: ||| |||||
2825 ATGTCAACAAAAGGAAGAAAT..... 2846
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
|||||
2847 .....CAAGAAAGATGATGCTTAATATC 2870
88 HisProPheGluSerAsnLeuGluPheThrSerLeuIleThrProAs 104
|||
2871 AAGCCTGTACAGACA..... 2885
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProT 121
: ||| |||
2886 .....GTTAATATCACTGCAGGCTTTCTCTG 2910
121 hr.....GlnProValAsnIleProLeuTyrAlaSerHis 132
: |||
2911 TGGTGTGCAGAAAGATAGCCAGTGTGATAATGCCAATGTATATCAAA 2960
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHi 146
|||||
2961 GGAGGCTCTAGTTTGTCTATCATCTCAGTTCAGAGCAACGAACTGG 3010
146 sLeuSerProProAsp.....SerProGluGlnValA 157
|||
3011 ACTCATTAATCCAAATAAACATGACCTTTTACAAAACCCATATCGTATAC 3060
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
|||||
3061 CACCACCTTTTCCCATCAAGTCATTTGTTAAAACTAAATGTAAGAAAAT 3110
174 Val.....IleSerProAspGlyLeuLe 181
: |||
3111 CTGCTAGAGGAAACTTTGAGCAACATTCATATCACCTGAAGAGAAAT 3160
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
: |||
3161 GGGAAATGAGAACATTCCAAGTACAGTACAGCACAAATTAGCCGTAATAACA 3210
193 .....PheThrGluLeuAspPheValAsnPheAsnAsp... 203
|||
3211 TTAGAGAAAATGTTTTTAAAGAACGCCGCTCAAGCAATATTAAATGAAGTA 3260
204 ...SerAlaValGlySerIleGlyGlyAlaGluGluLeuLeuGlySerPr 219
|||||
3261 GGTTCAGTACTAATGAGTGGGCTCCAGTATTAAATGAAATAGGTTC... 3308
219 oLeuSerValAspAspValGluSerThrIleSerPheSer...GlyProS 235
|||
3309 ....AGTGATGAAAACATTCAGCAAGCAACTAGGTAGAAACAGAGGGCCAA 3354
235 er.....SerProGluThrSerGln 241
: |||
3355 AATTGAATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGCTATATAA 3404
242 SerSerIleIleGluSerSer.....ProGluLeuTyrLys..... 253
: |||
3405 CAAAGCTTCTCTGGAAGTAATGTGAAGCATCTCAATATAAAAGCAAGA 3454
254 .....ValIleSerThrSerIleAspAlaSerLysArgPheS 267
|||||
3455 ATATGAAGAAGTAGTTCAGACTGTTAATACAGAT.....TTCT 3492
267 erProTyrSerArgSerSerLysSerLysGlnSerValLysThrSerAsp 283
|||||
3493 CTCATATCTGATTTCAAGTAATCAAGTATAGAACGCCTATAGGGAAGTAGTCAT 3542

```

```

284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluHi 300
|||||
3543 GCATCT.....CAGGTTTGTTCTGAGACACCTGATGA 3574
300 sValIleMetGluHisLeuAspLysLysLysLysLysLysLysLysLys 316
: |||
3575 CCGTCTAGATGCTGTAATAAAGGAAGATAGTCTGCTGCTGCTGCTGCTG 3624
317 .....LysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGlu 330
|||||
3625 ACATTAGGAAGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3674
331 .....AlaGlnGlyI 334
|||||
3675 CTTAGCAGGAGCTCTAGCCCTTTCACCCATACACATTTGGCTCAGGGTTA 3724
334 eLysGlyGluGlnGluLeuGluGluLeuAsnThrLysLysLysThrL 351
: |||
3725 CCGAAGAGGGGCCAAGAAATTAGAGTCTCAGAAAGAACTTATCTAGTG 3774
351 ysValAspAspLeu 355
: |||
3775 AGGATGAAGAGCTT 3788

```

seq_name: N_Geneseq_36:V32452

seq_documentation_block:

ID V32452 standard; cDNA; 8982 BP.

AC V32452;

DT 12-OCT-1998 (first entry)

DE Human receptor tyrosine kinase LMR2_h cDNA.

KW Receptor tyrosine kinase; LMR2_h; human; signal transduction;

KW cancer; neurodegenerative disorder; gene therapy; ss.

OS Homo sapiens.

FH Key

FT CDS Location/Qualifiers

FT 313..4824

FT sig_peptide /*tag= a

FT 313..381

FT /*tag= b

FT mat_peptide 382..4821

FT /*tag= c

FT WO9822507-A2;

PN 28-MAY-1998.

PD 21-NOV-1997; U22526.

PE 22-NOV-1996; US-031675.

PR (SUGB-) SUGEN INC.

PA Joho KE, Plowman GD;

PI WPI; 98-312419/27.

DR P-PSDB; W48845.

DR New isolated receptor tyrosine kinase genes - which are expressed in

PT neuronal tissues and tumour cells, useful as targets for

PT neurodegenerative disorders or cancers

PS Claim 1; Fig 6D-H; 87pp; English.

CC This nucleic acid molecule (NAM) codes for LMR2_r (see W48845), a

CC novel human receptor tyrosine kinase (RTK). A partial cDNA clone

CC was isolated from human heart cDNA using primers (see V42006 and

CC V42010) based on rat LMR1_r cDNA (see V32448). 2 LMR2_h cDNA

CC clones of 4349 and 5482 bp were isolated from an NCI-H460 human

CC lung carcinoma cell line cDNA library and from a SNB75 library,

CC respectively, using this partial clone. These clones span the

CC complete 8982 bp human LMR2 cDNA. NAMS (see V32448-56) coding for

CC novel kinases LMR1, LMR2 and LMR3 (see W48841-49) have been obtained

CC from rat, human and mouse sources. Expression of LMR1 and LMR3 is

CC highly restricted to neuronal tissues with minimal expression in

CC other adult or embryonic organs or in human tumour cell lines.

CC LMR2 expression is limited to adult neuronal tissues, but is also

CC very abundantly expressed in other non-neuronal foetal tissues and

CC in numerous tumour cell lines. Based on restricted expression of

CC all 3 LMRs to adult neuronal tissues and the up-regulation of LMR2

CC in a wide variety of tumour cell lines, these proteins may be

CC critical targets for neurodegenerative disorders or cancer. Claimed

CC probes and antibodies may be used for detecting neurodegenerative

FT
FT
FT


```

3621 CTTAGCAGGAGCTCTAGCCCTTCCACCATACACATTGGCTCAGGGTTA 3670
|||||
334 eLysGlyGluGlnGluGluGluGluLeuAsnThrLysLysLysThrL 351
::: |||||
3671 CCGAGAGGGGCCAAGAAATAGAGTCTCAGAGAGAACTTATCTAGTG 3720
|||
351 ysValAspAspLeu 355
::: |||||
3721 AGGATGAAGAGCTT 3734

```

seq_name: N_Geneseq_36.V60560

seq_documentation_block:

ID V60560 standard; cDNA; 5656 BP.
 AC V60560;
 DT 09-DEC-1998 (first entry)
 DE Tumorigenic BRCAL allele #5803.
 KW Tumorigenic allele; ovarian; breast cancer susceptibility gene; BRCAL;
 KW identification; specific mutation; allele specific antibody; detection;
 KW binding assay; treatment; ds.
 OS Homo sapiens.

Key Location/Qualifiers

FT CDS 120..200
 /*tag= a

US821328-A.

PN 13-OCT-1998.

PF 02-APR-1997; 825886.

PR 08-DEC-1993; US-163959.

PR 18-APR-1994; US-232535.

PR 20-OCT-1994; US-326983.

PR 19-APR-1995; US-425061.

PR 02-APR-1997; US-825886.

PA (REGC) UNIV CALIFORNIA.

PI Friedman L, King M, Lee M, Lynch E, Ostermeyer B,

PI Rowell S, Szabo C;

DR P-PSDB; W79886.

PT Poly(peptide(s) based on mutated BRCAL allelic sequences - useful
 for identification of specific allelic mutation(s) of the gene

PT causing breast cancer

PS Example 2; Columns 11-18; 62pp; English.

CC The present sequence represents a tumorigenic allele of the breast and
 CC ovarian cancer susceptibility gene (BRCAL). Exon 3 has been deleted
 CC in the present allele, which results in a truncated protein. Several
 CC other mutagenic alleles have also been identified (see V60560-71). The
 CC polypeptides are useful for the identification of specific mutations
 CC of BRCAL. The alleles are specific mutations of the C-terminal end of
 CC BRCAL, and can be used to compare with the translation product from a
 CC patient to identify the mutation that is causing the disease. In this
 CC respect, they can also be used to raise allele specific antibodies.
 CC They can also be used for detection purposes. The antibodies can be
 CC used in binding assays e.g. ELISA. The polypeptides can also be used
 CC for the treatment of breast cancer.

SQ Sequence 5656 BP; 1939 A; 1087 C; 1263 G; 1367 T;

alignment_scores:

Quality: 118.50 Length: 438
 Ratio: 0.605 Gaps: 20
 Percent Similarity: 44.749 Percent Identity: 19.178

alignment_block:

US-08-656-811A-1 x V60560 ..

Align seg 1/1 to: V60560 from: 1 to: 5656

13 AlaArgGluTrpGlyLeuGluMetProValValGlnThrAspGlyGlnPh 29

::: ||||| :::: |||||

2583 AGTCGGGAACACAGCATAGAATGCGAAGAGTGAACCTTGATGTCAGTA 2632

29 eGlyAspLeuLysSerThr.....

|||||:::||||

```

2633 T.....TTGCAGAAATACATTCAAGGTTTCAAAGCGCCAGTCATTGCTC 2676
36 .....SerArg 37
2677 CGTTTCAAATCCAGGAATCCAGAAGAGGAATGCAACATTTCTCTGCC 2726
38 HisGlyGlyAspGlySerLeuSerLeuGlnProGlnGlyAlaThrLeuLy 54
|||||::: |||
2727 CACTCTGG.....TCCTTAAAGAAACAAAGTCCAAAGTCACCTTTGA 2770
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
: ||| |||||
2771 ATGTGAACAAAAGGAAGAAAAT..... 2792
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
|||||::: |||
2793 .....CAGGAAGAAATGAGTCTAATATC 2816
88 HisProPheGlySerAsnLeuGluPheThrSerLeuLeuThrProAs 104
|||::: |||
2817 AAGCCTGTACAGACA..... 2831
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProT 121
::: ||| |||||
2832 .....GTTAATATCACTGCAGGCTTCTCTG 2856
121 hr.....GlnProValAsnIleProLeuTyAlaSerHis 132
::: |||||::: |||
2857 TGGTTGGTCAGAAAGATAGCCAGTTGATAATGCCAAATGTAGTATCAAA 2906
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHi 146
|||||::: |||
2907 GGAGCCTCTAGTTTGTCTATCATCTCAGTTCAGAGCAACAGAACTGG 2956
146 sLeuSerProProAsp.....SerProGluGlnVala 157
||| |||||
2957 ACTCATTACTCCAATAAACATGACGCTTTTACAAAACCCATATCGTATAC 3006
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
|||||::: |||
3007 CACCACCTTTTCCCCTCAAGTCATTTGTTAAACTAAATGTAAGAAAAT 3056
174 ValIle.....SerProAspGlyLeuLe 181
::: |||||::: |||
3057 CTGCTAGAGGAAACTTTGAGGAACATTCATGTCCCTCGAACAGAAAT 3106
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
|||||::: |||
3107 GGAATGAGAACATTCGAGTACAGTACGAGCACAATTAGCCGTAATAACA 3156
193 .....PheThrGluLeuAspPheValAsnPheAsnAsp... 203
|||||::: |||
3157 TTAGAGAAATGTTTAAAGAACCCAGCTCAAGCAATATTAAATGAAGTA 3206
204 ..SerAlaValGlySerIleGlyAlaGluGluLeuLeuGlySerPr 219
|||||::: |||
3207 GGTTCACGACTAATGAAGTGGCTCCAGTATTAAATGAATAGGTTC... 3254
219 oLeuSerValAspAspValGluSerThrIleSerPheSer...GlyProS 235
|||::: |||
3255 ...AGTCATGAAACATTCAGAGAACTAGGTAGAAACACAGAGGGCCAA 3300
235 er.....SerProGluThrSerGln 241
::: |||||::: |||
3301 AATTGAATGCTATGCTTAGATTAGGGGTTTGGCAACCTGAGGTCTATAAA 3350
242 SerSerIleIleGluSerSer.....ProGluLeuTyLys..... 253
::: |||||::: |||
3351 CAAAGTCTTCTCGAAGTAATTTGATGACATCTCGAATAAAAAAGCAAGA 3400
254 .....ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
|||||::: |||
3401 ATATGAAGAAGTAGTTTCAGACTGTTAATACAGAT.....TTCT 3438

```


This Page Blank (uspto)

Command line parameters:

```

seq_documentation_block:
; Sequence 1, Application US/08721684C
; Patent No. 5854016
; GENERAL INFORMATION:
; APPLICANT: Keegan, Kathleen S.
; TITLE OF INVENTION: NO. 5854016el CREBa Isoform
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA

```

```
?  
? COMPUTER READABLE FORM:  
? ZIP: 60606  
? COUNTRY: USA  
?  
? MEDIUM TYPE: Floppy disk  
?  
? COMPUTER: IBM PC compatible  
? OPERATING SYSTEM: PC-DOS/MS-DOS  
? SOFTWARE: Patentin Release #1.0, Version #1.30  
? CURRENT ADDRESS: N.D.M.
```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/33487
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
INFORMATION FOR SEQ. ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 390 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 304..1866
US-08-721-684C-1

```
alignment_scores:      Length: 423
                        Quality: 132.00
                        Gaps: 22
                        Ratio: 0.635
                        Percent Similarity: 49.173
                        Percent Identity: 21.513
```

```
alignment_block:
US-08-656-811A-1 x US-08-721-684C-1
```

Align seq 1/1 to: US-08-721-684C-1 from: 1 to: 3190

19 GluMetProValValGlnThrAspGlyGlnPheGlyAspLeuIysSer⁷⁴h 35

325 GAGCAGAGGGTCCTGCAGTGGGACCGCAGCTCAGCCAGCTC

35 r5e0rdrHjcs|vC|vdsq||5ser|oeC|v drcf|v|v|v|oe 5c

367

C
+
t
t
E
C
r
C
c
t
t
t
C
C
F
C
C
C
t
t
t
t
t
t

— 10 —

2020年12月10日

Align seg 1/1 to: US-09-005-970-1 from: 1 to: 3190

```
19 GluMetProValValGlnThrAspGlyGlnPheGlyAspLeuLysSerTh 35
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
325 GAGCAGAGCGTCTCGCAGTGGGACCGCAAGCTGAGCGAGCTG..... 366
35 rSerArgHisGlyGlyAspGlySerLeuSerLeuGlnProGlnGlyAlaT 52
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
367 .TCAGAGCCGGGAGAGACTGAGGCCCTCATGTACACACGACTTCTCGG 415
52 hrLeuLysLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMet 68
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
416 AGCTC...CTAGACGAGTTTCCCAAGACGCTCTGGGT..... 450
69 GluSerSerAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisG1 85
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
451 .....CAGCTCCTGAGTGAC..... 465
85 uArgLeuHisProPheGluSerAsnLeuLeuGluPheThrSerLeuIleT 102
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
466 .....CCTTCCTCTCA..... 477
102 hrProAspSerThrValSerLysAspIleLeuSerSerThrLeuGln 118
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
478 .....GAGACAGCGAGTCAATCGAGGTGGAGCCATCT..... 510
119 PheProThrGlnProValAsnIleProLeuTyzAlaSerHisGlyAlaG1 135
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
511 ...CCAACATCACCGG.....CCTCTCATCCAGGCTGAACACAGCTA 551
135 uaspPheSerAlaGluThrGluPheGluAsn.....HisLeuSerP 149
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
552 CTCTCTGAGCGAGGAGCCCGGACTCAGTCACCATTTACCCATGCGGCTA 601
149 roProAspSer.....ProGluGlnVal..... 156
602 CCAGCCACAGCTTCAATGACGAGGAGGTGGAGTGCAAAAAGGTACCTG 651
157 .....AlaProValIleAsnLeuGluProValGluLeuTh 168
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
652 TCTACAGAGTTTCTCTCAGCTACCATCAAGAAAGACCAATC..... 693
168 rAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyGlyMetG 185
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
694 .....ACAGAGGAGCGCCCGGACTT..... 717
185 lLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsnPhe 201
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
718 ..GTCCCTTCTGTCACTCTGACCATCACGCCATTTCC...ACTCCTTT 762
202 Asn.....AspSerAlaValGlySerIleGlyAl 212
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
763 GAAAGAAAGAGTCCCTCTGATATGATGCTGGGGGAGCTCTCATG 812
212 aGluGluLeuLeuGlySer...ProLeuSerValAspValGluSerT 228
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
813 CCAGAGCGTTATTCTTAAGATTAAAGTGGAGCCCAAGAAAGTGGATCAGT 862
228 hrIleSerPheSerGlyProSerSerProGluThrSerGlnSerSerIle 244
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
863 TCTTAAACTTCTCC...CCGAAGAAGAGCTCGGTGGATCAACTGCACCTA 909
245 lLeuSerSerProGluLeuTyzValIleSerThrSerSerIleAs 261
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
910 CCACCAACACACCCAGTAGTACACAGCAGTACTCTGAGGCGAGCTTGAG 959
261 pAlaSerLysArgPheSerProTyzSerArgSer..... 273
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
960 CCCCACCCACGCGCTCATCCCTTACGCTGTCTAGCCCTGTCAGGCCACAGCCCTG 1009
274 .....LysSerLysGlnSerValLysThrSerAsp....Ala 284
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
```

```
1010 TCAGAGCCATGCCCGGGGCCCTCTGCTTGTCCACATCTCCTCTCCTC 1059
285 LysAlaProArgLysThrArg..... 291
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1060 ACAGCTCCACATAAGCTGCAGGATCGGGCCCTCGTCTCTGACAGAAGA 1109
292 .....ThrProAlaGlnProValProGluHisValI 302
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1110 GGAGAAGAGAGCCCTGTTGCCGAGGCTATCCCTTCCACCAANGCTGC 1159
302 leMet.....GluHisLeuAspLysLysAspArgLysLysLeu 314
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1160 CTCTGCAAAATCTGAGGAGAGCCCTGAAGAAATCCGGAAGAAATCAT 1209
315 GlnAsnLysAsnAlaIleArgTyzArgMetLysLysLysGlyGluAl 331
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1210 AGAATAAGATTCTCCCAAGAAAGCAGGAGAAAGAAAGAAATATCAT 1259
331 aGlnGlyIleLysGlyGluGlnGluLeuGluLeuAsnThrLysL 348
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1260 GGACAGCCTGGAGAAAAAGTGGAGTCTTGTCAACTGAGAACTGGAGC 1309
348 euLysThrLysValAspAspLeuGln.....Arg 357
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1310 TTCGGAAGAGGTGGAGTCTGGAGAACCAATAGGACTCTCTCTTCAG 1359
358 GluIleLysTyzMetLysAsnLeu.....MetGluAspValCy 370
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1360 CAACTTCAGAAAGCTTCAGACTTTGGTGATGGGGAGNGTCTCTCGAACCTG 1409
370 sLysAlaLysGlyIleGln 376
   |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1410 CAAAGTTAGTGGCACACAG 1428
seq_name: /cgn2_6/ptodata/2/lna/5D_COMB.seq:US-08-319-866-7
seq_documentation_block:
; Sequence 7, Application US/08319866
; Patent No. 5929223
; GENERAL INFORMATION:
; APPLICANT: Tully, Timothy P.
; APPLICANT: Yin, Jerry C.
; APPLICANT: Regulski, Michael
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF GENES
; TITLE OF INVENTION: ASSOCIATED WITH LONG-TERM MEMORY
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/319,866
; FILING DATE: 7-OCT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL94-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 7:
```



```
alignment_scores:
  Quality: 124.50      Length: 372
  Ratio: 0.688        Gaps: 17
  Percent Similarity: 48.656  Percent Identity: 22.043

alignment_block:
US-08-656-811A-1 x US-08-993-118-1 ..
Align seg 1/1 to: US-08-993-118-1 from: 1 to: 4031

5 LeuSerPheGluAspPheGlnLeuAlaArgGluTrpGlyLeuGluMetPr 21
  ||| :|||:||||| :||| :|||:||||| :|||:||||| :|||:|||||
1892 CTTGAGGGGAGGAGATTCCAGTCTCT .....CTCCAGAGCCC 1929

21 oValValGlnThrAspGlyGlnPheGlyAspLeuLysSerThrSerArgH 38
  |||| :|||:||||| :||| :|||:||||| :|||:||||| :|||:|||||
1930 TGTGAGCATCTGCTCC .....TCTCCACTCCATCCAGTCTTCCCC 1970

38 IsGlyGlyAspGluSerLeuSerLeuGlnProGlnGlyAlaThrLeuLys 54
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1971 AGAGTTTCCCTGAGAGTTCTCAGAGTCTCCTGAGGGGCGCTGTC ..... 2014

55 LeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerSe 71
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2015 CAGTCTCTCTCCATAGTCTCCAGAGCCCTCTGAGGGGAGTCACTCCCA 2064

71 rAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeuH 88
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2065 ATCTCTCTCCAGAGTCTGAGAGTGCTCTGAGGGGAGGATTCCTGT 2114

88 IsProPheGluSerAsnLeuLeuGluPheThrSerLeuIleThrProAsp 104
  |||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2115 CTCCTCTCCAA .....ATTCTCTCAG 2134

105 AspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPhePro... 120
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2135 AGTCTCTTGAGGGGAGAGACTCCCTGCTCTCTCCACTTTCCTCAGT 2181

121 .....ThrGlnProValAsnIleProLeuT 129
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2182 GAGTCTCTGAGTGGGAGGACTCCCTCTCTCTCTCCACTTTCCTCAGT 2231

129 yrAlaSerHisGlyAlaGluAspPhe .....G1 144
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2232 TTCTCTCTCAGGGG...GAGGACTTCCAGTCTTCTCTCCAGAGTCTGTG 2278

138 .....SerAlaGluThrGluPhe .....G1 144
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2279 AGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2328

144 uAsnHisLeuSerProProAspSerProGluGlnValAlaProValIleA 161
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2329 GAGTCTCTCAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2371

161 snLeuGluProValGlu .....Leu 167
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2372 ..CAGAGACCTGTACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2419

168 ThrAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyGlyMe 184
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2420 CAAGTTCCTCAGAGAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2455

184 tGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsn 201
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2455 ..... 2455

201 heAsnAspSerAlaValGlySerIleGlyAlaGluGluLeuGly 217
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2456 .....GGGCTGCGCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2479

218 SerProLeuSer .....ValAspAspValGluSerThrIleSerPheSe 232
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2480 AGTCTGTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2529
```

```
232 rGlyPro.....SerSerProGluThrSerGlnSerSerIleIleGluS 247
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2530 TTCTCTGTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAAGA 2579

247 erSerProGluLeuThrLysValIleSerThrSerSerIleAspAlaSer 263
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2580 GTTCCCTGAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 2629

264 LysArgPheSerProTyrSerArgSerSer..... 273
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2630 ACTTCTCTGAGCCCATTCAGTCAAGAGTCCAGCCAGTCTCTCTCTCTCT 2679

274 .....LysSerLysGlnSerValLysThrSerAspAlaLysAlaProArgL 289
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2680 TACAAGTCTCTCAGACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2729

289 ysThrArgThrProAlaGlnProValProGluHisValIleMetGluHis 305
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2730 AGTCTCTGATGAGAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2779

306 LeuAspLysLysAspArgLysLysLeuGlnAsnLysAsnAlaIleAr 322
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2780 GTGAGAGAGTTCGCGGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2811

322 gTyrArgMetLysLys 327
  :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2812 ATATCAAGTGAAGCAG 2827

seq_name: /cgn2_6/ptodata/2/ina/6_COMB.seq:US-08-845-528C-1

seq_documentation_block:
; Sequence 1, Application US/08845528C
; Patent No. 6027924
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie;
; APPLICANT: DE SMET, Charles;
; APPLICANT: BOON-FALLEUR, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,528C
; FILING DATE: April 25, 1997
; CLASSIFICATION: 4335
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 638-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4031 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
; US-08-845-528C-1
```



```

; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
US-08-993-118-9

alignment_scores:
  Quality: 124.50      Length: 372
  Ratio: 0.688        Gaps: 17
  Percent Similarity: 48.656      Percent Identity: 22.043

alignment_block:
US-08-656-811A-1 x US-08-993-118-9 ..
Align seg 1/1 to: US-08-993-118-9 from: 1 to: 4225

5 LeuTrpSerGluAspPheGlnLeuAlaArgGluTrpGlyLeuGluMetPr 21
||| .....||| .....||| .....||| .....||| .....|||
2102 CTTGAGGGGAGGAAATTCAGTCTTCT.....CTCCAGAGCCC 2139

21 oValValGlnThrAspGlyGlnPheGlyAspLeuLysSerThrSerArgH 38
||| .....||| .....||| .....||| .....||| .....|||
2140 TGTGAGCATCTGCTCC.....TCTCCACTCCATCCAGTCTTCCCC 2180

38 isGlyGlyAspGluSerLeuGlnProGlnGlyAlaThrLeuLys 54
||||| .....||| .....||| .....||| .....||| .....|||
2181 AGAGTTTCCCTGAGAGTCTCAGAGTCTCTCTGAGGGGCTGTC..... 2224

55 LeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerSe 71
||||| .....||| .....||| .....||| .....||| .....|||
2225 CAGTCTCTCTCCATAGTCTCAGAGCCCTCTGAGGGGATGACCTCCCA 2274

71 rAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeuH 88
||||| .....||| .....||| .....||| .....||| .....|||
2275 ATCTCTCTCCAGAGTCTGAGAGTCTCTGAGGGGAGGATTCCTGT 2324

88 isProPheGluSerAsnLeuGluPheThrSerLeuIleThrProAsp 104
||||| .....||| .....||| .....||| .....||| .....|||
2325 CTCCTCTCCAA.....ATTCTCTCAG 2344

105 AspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPhePro.. 120
||| .....||| .....||| .....||| .....||| .....|||
2345 AGTCTCTTGAGGAGAGGACTCCTCTCTCT.....CTCCATTTTCTCTCA 2391

121 .....ThrGlnProValAsnIleProLeuT 129
||||| .....||| .....||| .....||| .....||| .....|||
2392 GAGTCTCTCAGTGGGAGGACTCCTCTCTCTCTCTCCACTTTTCTCAGT 2441

129 yrAlaSerHisGlyAlaGluAspPhe..... 137
||| .....||| .....||| .....||| .....||| .....|||
2442 TTCTCTCTCAGGGG...GAGGACTTCCAGTCTCTCTCTCCAGAGTCTGTG 2488

138 .....SerAlaGluThrGluPhe.....G1 144
||||| .....||| .....||| .....||| .....||| .....|||
2489 AGTATCTGCTCTCTCCACTCTTTGAGTCTCCCCAGAGTTTCCCTGA 2538

144 uAsnHisLeuSerProAspSerProGluGlnValAlaProValIleA 161
||| .....||| .....||| .....||| .....||| .....|||
2539 GAGTCTCTCAGAGTCTCTCTGAGGGGCTGCTCAG...TCTCTCTC... 2581

161 snLeuGluProValGlu.....Leu 167
||||| .....||| .....||| .....||| .....||| .....|||
2582 ..CAGAGACTGTGCTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 2629

168 ThrAlaSerHisMetThrValIleSerProAspGlyLeuGlyGlyMe 184
||||| .....||| .....||| .....||| .....||| .....|||
2630 CAAGTTTCCCATGAGTCTCAGAGTCTCTCTCTCTCTCTCTCTCTCTCT 2665

184 tGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsnP 201
||||| .....||| .....||| .....||| .....||| .....|||
2665 ..... 2665

201 heAsnAspSerAlaValcylserIleGlyGlyAlaGluLeuLeuGly 217
||| .....||| .....||| .....||| .....||| .....|||

```



```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-480-784-21

alignment_scores:
  Quality: 118.50      Length: 438
  Ratio: 0.605        Gaps: 20
Percent Similarity: 44.749 Percent Identity: 19.178

alignment_block:
US-08-656-811A-1 x US-08-480-784-21 ..
Align seg 1/1, to: US-08-480-784-21 from: 1 to: 4249

13 AlaArgGluTrpGlyLeuGluMetProValValGlnThrAspGlyGlnPh 29
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2048 AGTCGGAAACAAGCAGTAGAATGGAAGAAAGTGAACCTTGATGCTCAGTA 2097
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
29 eGlyAspLeuLysSerThr..... 35
   :  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2098 T.....TTGCAGATAACATTCAGGTTTCAAAGGCCGCCAGTCATTGCTC 2141
   36 .....  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2142 CGTTTTCAAATCCAGAAATGCAGAGAGGAATGCAACATTCCTGCC 2191
   38 HisGlyGlyAspGluSerLeuSerLeuGlnProGlnGlyAlaThrLeuLys 54
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2192 CACTCTGGG.....TCCTTAAAGAAACAAGTCCAAAGATCACITTTGA 2235
   54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
   :  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2236 ATGTCGAACAAAGGAAGAAAT..... 2257
   71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2258 .....  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
   88 HisProPheGluSerAsnLeuGluPheThrSerLeuIleThrProAs 104
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2282 AGCCTGTACAGACA..... 2296
   104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProT 121
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

2297 .....GTTAATATCACTGCAGGCTTTCCTCG 2321
121 hr.....GlnProValAsnIleProLeuTyrAlaSerHis 132
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2322 TGGTTGGTCAGAAAGATAAGCCAGTTGATAATGCCAAATGTAGTATCAAA 2371
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnH 146
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2372 GGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAGAGGCACGNAAGTGG 2421
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
146 sLeuSerProProAsp.....SerProGluGlnVala 157
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2422 ACTCATTACTCCAATAAACAATGGACTTTTACAAAACCCCATATCGTATAC 2471
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2472 CACCACCTTTTCCCATCAAGTCATTGTTTAAAACTAAATGTAAGAAAAAT 2521
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
174 ValIle.....SerProAspGlyLeuLe 181
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2522 CTGCTAGAGGAAAACTTTGAGGAACATTCAAATGTCACCTGAAGAGAAAT 2571
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2572 GGAATAAGAAACATTCCAAGTACACTGAGCACAAATTACCCGTAATAACA 2621
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
193 .....PheThrGluLeuAspPheValAsnPheAsnAsp... 203
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2622 TTAGAGAAAATGTTTTTAAAGAACCCAGCTCAAGCAATATTATGAAGTA 2671
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
204 ...SerAlaValGlySerIleGlyAlaGluGluLeuLeuGlySerPr 219
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2672 GGTTCAGTACTAATGAAGTGGCTCCAGTATTAAATGAATAGGTTCC... 2719
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
219 oLeuSerValAspAspValGluSerThrIleSerPheSer...GlyProS 235
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2720 ....AGTGATGAAGAACATTCACAGCAGACTAGGTAGAAACAGAGGCCCAA 2765
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
235 er.....SerProGluThrSerGln 241
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2766 AATTGAATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGCTCTATAA 2815
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
242 SerSerIleGluSerSer.....ProGluLeuTyrLys..... 253
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2816 CAAAGCTTCCTCGGAAGTAAATTGTAAGCATCCTGAAATAAAAAAGCAAGA 2865
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
254 .....ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2866 ATATGAAGAAGTAGTTACAGACTGTTAATACAGAT.....TTCT 2903
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
267 erProTyrSerArgSerLysSerLysGlnSerValLysThrSerAsp 283
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2904 CTCCATATCTGATTTCAGATAACTTAGAACACGCTATGGAAAGTAGTCAT 2953
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluH 300
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2954 GCATCT.....CAGGTTTGTCTGTGAGACACCTGATGA 2985
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
300 sValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn.. 316
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2986 CTTGTAGATGATGGTGAATAAAGGAAGATACTAGTTTTTGTCTGAAAATG 3035
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
317 .....LysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGlu 330
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3036 ACATTAAAGAAAGTCTCTGCTGTTTTTAGCAAAAGCGCTCCAGAAAGGAG 3085
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
331 .....AlaGlnGlyI 334
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3086 CTAGCAGAGTCCCTAGCCCTTTCACCCATACACATTTGGCTCAGGGTTA 3135
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
334 eLysGlyGluGlnGluLeuGluLeuAsnThrLysLysLeuLysThrL 351
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3136 CCGAAGAGGGGCCAAGAAATTAGAGTCCCTCAGAAAGAACTTATCTAGTG 3185
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```
351 ysValAspLeu 355
:: : : : :
3186 AGGATGAAGAGCTT 3199

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq.us-08-483-553-21
seq_documentation_block:
; Sequence 21, Application US/08483553
; Patent No. 5709999
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,553
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
```

```
US-08-483-553-21

alignment_scores:
  Quality: 118.50      Length: 438
  Ratio: 0.605        Gaps: 20
  Percent Similarity: 44.749  Percent Identity: 19.178

alignment_block:
US-08-656-811A-1 x US-08-483-553-21

Align seg 1/1 to: US-08-483-553-21 from: 1 to: 4249

13 AlaArgGluTrpGlyLeuMetProValValGlnThrAspGlyGlnPh 29
:::||||| : : : : :
2048 AGTCGGGAACAAGCATAGAAATGGAACAAAGTGAACCTTGATGCTCAGTA 2097
:::||||| : : : : :
29 eGlyAspLeuLysSerThr..... 35
: |||: : : : :
2098 T.....TTGCAGATACATTCAAGGTTTCAAAGCGCCAGTCATTGCTC 2141
36 .....SerArg 37
|||
2142 CGTTTCAAATCCAGGAATGCAGAAGAGGAATGTGCAACATTCTCTGCC 2191
38 HisGlyGlyAspGluSerLeuGlnProGlnGlyAlaThrLeuLy 54
|||||: : : : :
2192 CACCTGGG.....TCCTTAAAGAACAAAGTCCAAAGTCACCTTTGA 2235
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSers 71
: ||| : : : : :
2236 ATGTGAACAAAAAGGAAGAAAT..... 2257
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
||| : : : : :
2258 .....CAAGGAAGAATGAGTCTAATATC 2281
88 HisProPheGluSerAsnLeuGluPheThrSerLeuLeuThrProAs 104
||| : : : : :
2282 AAGCCTGTACAGACA..... 2296
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProt 121
: : : : : ||| : : : : :
2297 .....GTTAATATCATCATCGAGGCTTCTCTG 2321
121 hr.....GlnProValAsnIleProLeuTyrAlaSerHis 132
: : : : :
2322 TGGTGGTCAGAAAGATAGCCAGTTGATATGCCAAATGTAGTATCAAA 2371
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHI 146
|||: : : : : ||| : : : : :
2372 GGAGGCTCTAGTTTGTCTATCATCTCAGTTCAGAGCAACGAACTGG 2421
146 sLeuSerProProAsp.....SerProGluGlnVala 157
||| : : : : :
2422 ACTCATTTACTCCAAATAAATGAGGCTTTTACAAAACCCATATCGTATAC 2471
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
|||: : : : : ||| : : : : :
2472 CACCACCTTTTCCCATCAAGTCATTTGTTAAAACTAAATGTAGAGAAAAT 2521
174 Vallle.....SerProAspGlyLeuLe 181
: : : : : |||: : : : :
2522 CTGCTAGAGAGAAACTTTGAGGAACATTCATGTCCCTGAAGAGAAAT 2571
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
|||: : : : : ||| : : : : :
2572 GGGAAATGAGAACATTCACAGTACAGTACAGTACAGTACAGTACAGTACAGT 2621
193 .....PhethrGluLeuAspPheValAsnPhAspAsp... 203
||| : : : : : |||: : : : :
2622 TTAGAGAAAATGTTTTTAAAGAACCCAGCTCAAGCAATATTAAATGAGTA 2671
```



```
2258 .....CAAGGAAACAATGAGTCTAATATC 2281
88 HisProPheGluSerAsnLeuGluPheThrSerLeuIleThrProAs 104
   ||| :|:|:|
2282 ANGCTGTACAGACA..... 2296
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProT 121
   :|:|:| :|:|:| :|:|:|
2297 .....GTTAATATCACTGCAGGCTTTCCTG 2321
121 hr.....GlnProValAsnIleProLeuTyrAlaSerHis 132
2322 TGGTTGGTCAGAAAGATAGCCAGTTGTAATGCCAAATGTATATCAA 2371
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHi 146
   |||:|:|:| ||| :|:|:|:|
2372 GGAGGCTCTAGTGTTCATCATCTCAGTTCAGAGCAACGAACCTGG 2421
146 sLeuSerProProAsp.....SerProGluGlnValA 157
   |||:|:|:| |||:|:|:|
2422 ACTCATTAATCCAAATAACATGGACTTTTACAAAACCCATATCGTATAC 2471
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
   |||:|:|:| :|:|:|:| |||
2472 CACCACTTTTCCCATCAAGTCATTGTTAAAACTAAATGTAAAGAAAAT 2521
174 ValIle.....SerProAspGlyLeuLe 181
   :|:|:|:| |||:|:|:|
2522 CTGCTAGAGGAAACTTTGAGGACATTCATGTCCCTGAAAGAAAT 2571
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
   |||:|:|:| :|:|:| |||
2572 GGGAAATGAGAACATTCCAAGTACAGTGAGCACCAATTAGCCGTAATAACA 2621
193 .....PheThrGluLeuAspPheValAsnPheAsnAsp... 203
   |||:|:|:| ||| ||| :|:|:|:|
2622 TTAGAGAAATGTTTAAAGAGCAGCTCAAGCAATATTAAATGAAGTA 2671
204 ...SerAlaValGlySerIleGlyAlaGluLeuGlySerPr 219
   |||:|:|:| :|:|:|:| |||:|:|
2672 GTTCCAGTACTAATGAGTGGCTCCAGTATTATGAATAGGTTC... 2719
219 oLeuSerValAspValGluSerThrIleSerPheSer...GlyProS 235
   ||| :|:|:|:| :|:|:|:|
2720 ...AGTGATGAACAATTCAGACGAACTAGGTAGAAAACAGAGGCCAA 2765
235 er.....SerProGluThrSerGln 241
   :|:|:|:| |||:|:|:|
2766 AATCGAATGCTATGCTTAGATTAGGGTTTTCACACCTGAGGTCTATAAA 2815
242 SerSerIleIleGluSerSer.....ProGluLeuTyrLys..... 253
   :|:|:|:| |||:|:|:|
2816 CAAAGTCTCTCGGAAGTAATGTAGGATCTCTGAATATAAAMACAGA 2865
254 .....ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
   |||:|:|:| ||| :|:|:|:|
2866 ATATCAAGAAGTAGTTCAGACTGTTAATACAGAT.....TTCT 2903
267 erProTyrSerArgSerLysSerLysGlnSerValLysThrSerAsp 283
   |||:|:|:| |||:|:|:|
2904 CTCCATATCTGATTTCAGATAACTTTAGAACACCCCTATGGGAAGTAGTCAT 2953
284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluH 300
   |||:|:|:| |||:|:|:|
2954 GCATCT.....CAGGTTGTCTGAGACACCTGATGA 2985
300 sValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn. 316
   :|:|:|:| |||:|:|:|
2986 CTGTTAGATGATGCTGAAATAAGGAAGATAGTCTTTGCTGAAATG 3035
317 .....LysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGlu 330
   :|:|:|:| |||:|:|:|
3036 ACATTAAAGAAAGTTCTGCTGTTTTCAGAAAGCGTCCAGAAAGGAGAG 3085
```

```
331 .....AlaGlnGlyI 334
   |||:|:|:|
3086 CTTACGAGAGTCTAGCCCTTTCACCCATACACATTTGGCTCAGGGTTA 3135
334 eLysGlyGluGluGlnGluLeuGluGluLeuAsnThrLysLeuLysThrL 351
   :|:|:|:| :|:|:|:| :|:|:|:|
3136 CGGAAGAGGGCCAAAGAAATTAGAGTCTCAGAGAACTTATCTACTG 3185
351 ysValAspAspLeu 355
   :|:|:|:| |||
3186 AGGATGAAGAGCTT 3199
seq_name: /cgn2_6/ptodata/2/ina/5b_COMB.seq:US-08-483-5548-21
seq_documentation_block:
; Sequence 21, Application US/08483554B
; Patent No. 5747282
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,554B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4249 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-483-554B-21

alignment_scores:
  Quality: 118.50      Length: 438
  Ratio: 0.605        Gaps: 20
Percent Similarity: 44.749 Percent Identity: 19.178

alignment_block:
US-08-656-811A-1 x US-08-483-554B-21 ..
Align seg 1/1 to: US-08-483-554B-21 from: 1 to: 4249

13 AlaArgGluTrpGlyLeuGluMetProValGlnThrAspGlyGlnPh 29
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2048 AGTCGGGAAACAAAGCATAGAAATGGAAGAAAGTGAACCTTGATCCTCAGTA 2097
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
29 eGlyAspLeuLysSerThr..... 35
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2098 T.....TTGCAGATACATTCAGAGTTTCAAGGTTTCAAGGCCAGTCATTGGCTC 2141
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
36 .....SerArg 37
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2142 CGTTTTCAAATCCAGGAATGCGAAGAGGAATGTCACATCTCTGCC 2191
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
38 HisGlyGlyAspGluSerLeuSerLeuGlnProGlnGlyAlaThrLeuLy 54
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2192 CACTCTGGG.....TCCTTAAGAAACAAAGTCCAAAGTCACTTTTGA 2235
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
54 sLeuGlyProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2236 ATGTGAACAAAGGAAGAAAT..... 2257
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2258 .....CAAGGAAGAAATGAGTCTAATATC 2281
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
88 HisProPheGluSerAsnLeuLeuGluPheThrSerLeuIleThrProAs 104
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2282 AAGCCTGTACAGACA..... 2296
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProT 121
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2297 .....GTTAATATCACTGCAGGCTTTCCTG 2321
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
121 hr.....GlnProValAsnIleProLeuTyrrAlaSerHis 132
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2322 TGGTTGGTCAGAAAGATAAGCCAGTTGTAATGCCAAATGTAGTATCAAA 2371
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHi 146
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2372 GGAGGCTCTAGGTTTGTCTATCATCTCAGTTCAGAGGCAACGAACTGG 2421
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
146 sLeuSerProProAsp.....SerProGluGlnVala 157
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2422 ACTCATTTACTCCAAATAAATACATGGAGCTTTTACAAAACCCATATCGTATAC 2471
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2472 CACCACCTTTTCCCATCAAGTCATTGTTTAAAACTAAATGTAAGAAAAAT 2521
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
174 ValIle.....SerProAspGlyLeuLe 181
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2522 CTGCTAGAGGAAAACTTTGAGGNACATTCATGTCACCTGAAGAGAAAT 2571
```

```
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2572 GGGAAATGAGAACATTCCAAGTACAGTGAGCACAATTAGCCGTAATAACA 2621
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
193 .....PheThrGluLeuAspPheValAsnPheAsnAsp..... 203
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2622 TTAGAGAAATGTTTTTAAAGAAGCCAGCTCAAGCANTATTAATGANGTA 2671
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
204 ...SerAlaValGlySerIleGlyAlaGluGluLeuLeuGlySerPr 219
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2672 GGTTCCAGTACTAATGAAGTGGCTCCAGTATTAAATCAAAATAGTTCC... 2719
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
219 oLeuSerValAspAspValGluSerThrIleSerPheSer...GlyProS 235
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2720 ...ACTGATGAAACATTCACAGCAACTAGGTAGAAACAGAGGCCAA 2765
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
235 er.....SerProGluThrSerGln 241
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2766 AATTGAATGCTATGCTAGATTAGGGGTTTTGCAACCTGAGGTCTATATA 2815
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
242 SerSerIleIleGluSerSer.....ProGluLeuTyrrLys..... 253
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2816 CAAAGTCTCCCTGGAAGTAATTCCTAAGCATCTCTGAAATAAAAGCAAGA 2865
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
254 .....ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2866 ATATGAAGAAGTAGTTTCAGACTGTTAATACAGAT.....TTCT 2903
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
267 erProTyrrSerArgSerSerLysSerGlnSerValLysThrSerAsp 283
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2904 CTCATATCTGATTTCAGATAACTTAGACAGACGCCTATGGGAAGTAGTCAT 2953
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluH 300
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2954 GCATCT.....CAGGTTTGTCTGACACACCTGATGA 2985
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
300 sValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn. 316
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2986 CCTGTTAGATGATGCTGCTGCTTTTAGCAAAAGCGTCCAGAAAGGAG 3035
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
317 .....LysAsnAlaAlaIleArgTyrrArgMetLysLysGlyGlu 330
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3036 ACATTAGGAAGATTCTGCTGCTTTTAGCAAAAGCGTCCAGAAAGGAG 3085
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
331 .....AlaGlnGlyI 334
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3086 CTTAGCAGGAGTCTAGCCCTTTCACCCATACACATTTGCTCAGGGTTA 3135
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
334 eLysGlyGluGluGlnGluLeuGluLeuAsnThrLysLysLeuThrL 351
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3136 CCGAAGAGGGGCCAAGAAATTAGAGTCTCTCAGAAGAGAACTTATCTAGTG 3185
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
351 ysValAspAspLeu 355
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
3186 AGGATGAAGAGCTT 3199
   :|||:||||| :|||:||||| :|||:||||| :|||:|||||
```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-488-011B-21

seq_documentation_block:
; Sequence 21, Application US/08488011B
; Patent No. 5753441
; GENERAL INFORMATION:
; APPLICANT: SKolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer

2904 CTCATATCTGATTTCAGATACTTAGACACGCCTATGGAGTAGTCAT 2953
284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluHi 300
2954 GCAATCT.....CAGGTTTCTTGAGACACCTGATGA 2985
300 sValileMetGluHisLeuAspLysLysAspArgLysLysLysGluAsn. 316
2986 CCGTGTAGTATGCTGAAATAGGAAGATAGTCTTTGCTGAAATG 3035
317LysAsnAlaLalleArgTyrArgMetLysLysLysGlyGlu 330
3036 ACATTAAGGAAGTCTGCTGTTTTAGCAAAAGCGTCCAGAAAGGAGAG 3085
331
3086 CTTAGCAGGAGTCTTAGCCCTTCCACCATACACATTTGGCTCAGGGTTA 3135
334 eLysGlyGluGluGlnGluLeuGluGluLeuAsnThrLysLeuLysThrL 351
3136 CGGAAGAGGGCCAAAGAAATAGATGCTCCTCAGAGAGAACTTATCTAGTG 3185
351 ysValAspAspLeu 355
3186 AGGATGAAGAGCTT 3199
seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-10202-21
seq_documentation_block:
Sequence 21, Application PC/TUS9510202
GENERAL INFORMATION:
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Simard, Jacques
APPLICANT: Emi, Mitsuru
APPLICANT: Nakamura, Yusuke
APPLICANT: Dutcher, Francine
TITLE OF INVENTION: In vivo Mutations and Polymorphisms
TITLE OF INVENTION: In the 17q-Linked Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRES:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10202
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08-308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221

FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 4249 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
PCT-US95-10202-21
alignment_scores:
Quality: 118.50 Length: 438
Ratio: 0.605 Gaps: 20
Percent Similarity: 44.749 Percent Identity: 19.178
alignment_block:
US-08-656-811A-1 x PCT-US95-10202-21 ..
Align seg 1/1 to: PCT-US95-10202-21 from: 1 to: 4249
13 AlaArgGluTrpGlyLeuGluMetProValGlnThrAspGlyGlnPh 29
:::||||| :::::||||| :::::|||||
2048 ACTCGGAAACAACATAGACATAAATGGAAGAAAGTGAACCTTGATGCTCAGTA 2097-
29 eGlyAspLeuLysSerThr..... 35
: |||:::||||
2098 T.....TTGCAGATACATTCAAGGTTTCAAAAGCCAGTCATTTGCTC 2141
36SorArg 37
2142 CTTTTCAANTCCAGGAATGCAGAGAGGAATGCAACATTTCTCTGCC 2191
38 HisGlyGlyAspGluSerLeuSerLeuGlnProGlnGlyAlaThrLeuLys 54
||||||| :::::|||||
2192 CACTCTGGG.....TCCTTAAAGAAACAAGTCCAAAGTCACTTTTCA 2235
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSerS 71
: ||| :::::||||
2236 ATGTGAACAAAAGGAAGAAAT..... 2257
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
2258CAAGGAAGAAATGAGTCTAATATC 2281
88 HisProPheGluSerAsnLeuLeuGluPheThrSerLeuIleThrProAs 104
||| :::::
2282 AAGCCTGTACACACA..... 2296
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProt 121
:::|||||
2297GTTAATATCACTGCAGGCTTTCCCTG 2321
121 hr.....GlnProValAsnIleProLeuTyrAlaSerHis 132
:::|||||
2322 TGGTTGGTCAGAAAGATAGCCAGTTGATGATGCAAAATGTATATCAAA 2371
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnH1 146
||||::: ||| :::::|||||
2372 GGAGGCTAGGTTTGTCTATCATCTCAGTCAAGAGCGCAACGAAACTGG 2421
146 sLeuSerProProAsp.....SerProGluGlnValA 157

```

101
2422 ACTATTACTCAATAAATCAATGAGCTTTTACAAAACCCATATCGTATAC 2471
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
2472 CACCACTTTTCCCATCAAGTCACTTTGTGTAAACCTAAATGTAAGAAAAT 2521
174 ValIle.....SerProAspGlyLeuLe 181
2522 CTGCTAGAGGAAACTTTGAGGAACATTCATGTCACCTGAAAGAAAT 2571
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
2572 GGAATGAGAAACATTCACAGTACAGTACAGCACAATTAGCCGTAATAACA 2621
193 .....PheThrGluLeuAspPheValAsnPheAsnAsp... 203
2622 TTAGAGAAAATGTTTAAAGAACCCAGCTCAAGCAATTAATTAAGATA 2671
204 ...SerAlaValGlySerIleGlyAlaGluLeuLeuGlySerPr 219
2672 GGTTCACGACTAATGAGTGGCTCCAGTATTATGAATAGGTTC... 2719
219 oLeuSerValAspValGluSerThrIleSerPheSer...GlyProS 235
2720 ...AGTCATGAAAACATTCAGACGAGAACTAGGTAGAAAACAGAGGCCAA 2765
235 er.....SerProGluThrSerGln 241
2766 AATTGAATGCTATGCTTAGATTAGGGGTTTGCACCTGAGGTCTATAAA 2815
242 SerSerIleIleGluSerSer.....ProGluLeuTyrLys... 253
2816 CAAGCTCTCTCGNAGTAATGTGAAGCATCTGAAATAAAAACCAAGA 2865
254 .....ValIleSerThrSerIleAspAlaSerLysArgPheS 267
2866 ATATGAAGAAGTAGTTCAGACTGTTAATACAGAT.....TTCT 2903
267 erProTyrSerArgSerSerLysSerLysGlnSerValLysThrSerAsp 283
2904 CTCCATATCTGATTTCAGATAACTTAGAACACCCCTATGGGAAGTAGTCAT 2953
284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluHi 300
2954 GCATCT.....CAGTTTGTCTGAGACACCTGATGA 2985
300 sValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn. 316
2986 CCTGTTAGATGATGTTGTAATAAAGGAAGATAGTCTTGTGTAAGATG 3035
317 .....LysAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGlu 330
3036 ACATTAGGAAGTCTGCTGTTTTAGCAAAAGCGTCCAGAAAGGAGAG 3085
331 .....AlaGlnGlyI 334
3086 CTTAGCAGGAGTCTAGCCCTTTACCCATACACATTTGGCTCAGGGTTA 3135
334 eLysGlyGluGluGlnGluLeuGluLeuAsnThrLysLeuLysThrL 351
3136 CCGAAGAGGGGCCAAGAATAATAGATCTCTCAGAGAGAACTATCTAGTG 3185
351 ysValAspAspLeu 355
3186 AGGATGAAGAGCTT 3199
```

```

seq_name: /cgn2_6/ptodata/2/lna/PCTUS_COMB.seq:PCT-US95-10203-21
seq_documentation_block:
; Sequence 21, Application PC/TUS9510203
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
```

```

; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10203
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08-308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; PCT-US95-10203-21
```

```

alignment_scores:
Quality: 118.50 Length: 438
Ratio: 0.605 Gaps: 20
Percent Similarity: 44.749 Percent Identity: 19.178
alignment_block:
```



```
US-08-656-811A-1 x PCT-US95-10203-21 ..
Align seg 1/1 to: PCT-US95-10203-21 from: 1 to: 4249
13 AlaArgLuprPolpLeuGluMetProValGlnThrAspGlyGlnPh 29
   ::::::::::::::::::::
2048 AGTCGGGAAACAAGCATAGAAATGGAAGAAAGTGAACACTTGATGCTCAGTA 2097
   ::::::::::::::::::::
29 eGlyAspLeuLysSerThr.
   : ::::::::::::::
2098 T.....TTGCAGATACATTCAGAGTTTCAAGGCCAGCTCATTTGCTC 2141
   36 .....SerArg 37
2142 CGTTTCAATCCAGGAATGCGAAGAGGAATGTCACACATCTCTGCC 2191
38 HisGlyGlyAspGluSerLeuGlnProGlnGlyAlaThrLeuLys 54
   ::::::::::::::::::::
2192 CACTCTGGG.....TCCCTTAAAGAAACAAGTCCAAAGTCACTTTTGA 2235
54 sLeuGluProPheGluGluAspValLeuGlyAlaGluTrpMetGluSers 71
   : ::::::::::::::
2236 ATGTGAACAAAGGAAGAAAT..... 2257
71 erAspLeuGlySerPheLeuAspAlaLeuGlyAspAsnHisGluArgLeu 87
   : ::::::::::::::
2258 .....CAAGGAAGAATGAGTCTAATATC 2281
88 HisProPheGluSerAsnLeuLeuGluPheThrSerLeuIleThrProAs 104
   : ::::::::::::::
2282 AAGCCTGTACAGACA..... 2296
104 pAspSerThrValSerLysAspIleLeuSerSerThrLeuGlnPheProt 121
   : ::::::::::::::
2297 .....GTAAATATCACTGCAGGCTTCTCCTG 2321
121 hr.....GlnProValAsnIleProLeuThrAlaSerHis 132
   : ::::::::::::::
2322 TGGTTGGTCAGAAAGATAAGCCAGTGTAAATGCCAAATGTAGTATCAAA 2371
133 GlyAlaGluAspPheSerAlaGluThrGluPhe.....GluAsnHis 146
   : ::::::::::::::
2372 GGAGGCTCTAGTTGTCTATCATCTCAGTTCAGAGGCAAGCAAACTGG 2421
146 sLeuSerProProAsp.....SerProGluGlnValA 157
   : ::::::::::::::
2422 ACTCATTAATCCAAATAAATGAGTCTTTTACAAAACCCATATCGTATAC 2471
157 laProValIleAsnLeuGluProValGluLeuThrAlaSerHisMetThr 173
   : ::::::::::::::
2472 CACCACTTTTCCCACTCAAGTCTTTGTTTAAAACTAAATGTAGAAATAAT 2521
174 ValIle.....SerProAspGlyLeuLe 181
   : ::::::::::::::
2522 CTGCTAGAGGAAACCTTTGAGGAACATTCATCTCACCTGAAAGAGAAAT 2571
181 uGlyGlyMetGluLeuAlaSerGluSerLeuThr..... 192
   : ::::::::::::::
2572 GGAATAGAGAACATTCACCAAGTACAGTACAGTACAGTACAGTACAGTAC 2621
193 .....PheThrGluLeuAspPheValAsnPheAsnAsp... 203
   : ::::::::::::::
2622 TTAGAGAAATGTTTTTAAAGAGGCCAGCTCAAGCAATATTAATGAAGTA 2671
204 ...SerAlaValGlySerIleGlyGlyAlaGluLeuLeuGlySerPr 219
   : ::::::::::::::
2672 GGTTCAGTACTAATGAAGTGGCTCCAGTATTATTAATGAATAGTTCC... 2719
219 oLeuSerValAspAspValGluSerThrIleSerPheSer...GlyProS 235
   : ::::::::::::::
2720 ....ACTGATGAAACATTCACGACAGCACTAGTAGAACAAGAGGCGCAA 2765
235 er.....SerProGluThrSerGln 241
```

```

2766 AATTGAATGCTATGCTTAGATTAGGGCTTTTGCACACTGAGGCTCTATAAA 2815
   ::::::::::::::::::::
242 SerSerIleIleGluSer.....ProGluLeuTyrlLys..... 253
   : ::::::::::::::
2816 CAAAGTCTCTCTGGAAGTAATTCTAAGCATCCTGAAATAAAAAAGCAAGA 2865
   : ::::::::::::::
254 .....ValIleSerThrSerSerIleAspAlaSerLysArgPheS 267
   : ::::::::::::::
2866 ATATGAAGAAGTAGTACTGCTTAATACAGAT.....TTCT 2903
267 erProTyrlSerArgSerSerLysGlnSerValLysThrSerAsp 283
   : ::::::::::::::
2904 CTCATATCTGATTTCAGATAACTTACACAGCCTATGGGAAGTAGTCAT 2953
284 AlaLysAlaProArgLysThrArgThrProAlaGlnProValProGluH 300
   : ::::::::::::::
2954 GCATCT.....CAGGTTTGTCTCAGACACCTGATGA 2985
300 sValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsn. 316
   : ::::::::::::::
2986 CCTGTTAGATGATGGTGAATAAAGGAAGATACTAGTTTGTGCTGAAATG 3035
317 .....LysAsnAlaAlaIleArgTyrlArgMetLysLysGlyGlu 330
   : ::::::::::::::
3036 ACATTAAGGAAAGTTCTGCTGTTTGTAGCAAAAGCGTCCAGAAAGGAGAG 3085
331 .....AlaGlnGlyI 334
   : ::::::::::::::
3086 CTTAGCAGGAGCTCTAGCCCTTTCACCCCATACACATTTGCTCAGGGTTA 3135.
334 eLysGlyGluGluGlnGluLeuGluLeuAsnThrLysLysLeuLysThrL 351
   : ::::::::::::::
3136 CGAAGAGGGGCCAAGAAATTAGAGTCCCTCAGAAGAGAACTTATCTAGTG 3185
351 ysValAspAspLeu 355
   : ::::::::::::::
3186 AGGATCAAGAGCTT 3199
seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: PCT-US95-10220-21
seq_documentation_block:
; Sequence 21, Application PC/TUS9510220
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: Method for diagnosing a
; TITLE OF INVENTION: Predisposition for Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10220
; FILING DATE:
; CLASSIFICATION:
```


3086 CTTACGAGGATCCTAGCCCTTCACCCATACACATTGGCTCAGGGTTA 3135
|||||
334 eLysGluGluGlnGluLeuGluGlnLeuAsnThrLysLeuLysThrL 351
::: :::::|||||::: ::::: ::::: :::::
3136 CCGAGAGGGGCCAAGAAATTAGAGTCTCAGAAGAGAAGCTTATCTAGTC 3185

351 ysValAspLeu 355
:: :::::||||
3186 AGGATCAAGAGCTT 3199

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2000, 23:32:01 ; Search time 36.41 Seconds
(without alignments)
644.173 Million cell updates/sec

Title: US-08-656-811a-1
Perfect score: 1928
Sequence: 1 MELDLWSEDFQLAREWGLEM.....KYMKNLMEDVCKAKGIQLKM 379

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_64:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264	13.7	351	2 A45377	transcription fact
2	248	12.9	381	2 S26812	transcription fact
3	163	8.5	68	2 D34223	transcription fact
4	148.5	7.7	488	2 T27532	hypothetical prote
5	145	7.5	436	2 A37953	transcription regu
6	143.5	7.4	208	2 T24446	hypothetical prote
7	138.5	7.2	389	1 A39429	cAMP response elem
8	137	7.1	358	2 C42026	cyclic AMP respons
9	137	7.1	456	2 B42026	cyclic AMP respons
10	137	7.1	505	1 S05380	cAMP response elem
11	135	7.0	486	2 JC4028	activating transcr
12	134.5	7.0	330	2 T38374	probable bzip tran
13	134	7.0	647	2 S26386	transcription fact
14	133	6.9	448	2 A42026	cAMP response elem
15	133	6.8	483	2 S12741	transcription fact
16	130.5	6.8	582	2 S10099	transcription fact
17	129.5	6.7	313	2 A34785	DNA-binding protei
18	129.5	6.7	457	2 T21063	hypothetical prote
19	127	6.6	181	2 A54023	transcription fact
20	127	6.6	460	2 T19062	hypothetical prote
21	127	6.6	638	2 B35816	transcription regu
22	126.5	6.6	338	2 A53066	CCAAT enhancer-bin
23	126	6.5	699	2 T09069	probable cAMP-resp
24	126	6.5	938	2 A56731	chromatin assemb
25	124.5	6.5	2774	2 A43359	microtubule-associ
26	122.5	6.4	222	2 C34223	transcription fact
27	121	6.3	424	2 T10985	regulator protein
28	121	6.3	721	2 A33319	microtubule-associ
29	121	6.3	728	2 A81385	probable ATP /GTP

30	120.5	6.2	654	2 A34734	transcription fact
31	119	6.2	341	2 S67641	hypothetical prote
32	119	6.2	439	2 JC1178	transforming prote
33	119	6.2	853	2 T46347	hypothetical prote
34	119	6.2	1528	2 JS0703	DNA topoisomerase
35	118.5	6.1	590	2 S66956	hypothetical prote
36	118.5	6.1	1151	2 T24541	hypothetical prote
37	118.5	6.1	1863	1 A58881	breast/ovarian can
38	118	6.1	378	2 T12621	Dc3 promoter-bindi
39	118	6.1	547	2 T40314	hypothetical colle
40	118	6.1	551	2 S67788	hypothetical prote
41	117.5	6.1	427	2 S53021	G-box-binding prot
42	117.5	6.1	682	2 A42141	transcription fact
43	117	6.1	355	2 T40132	ATF/CREB-family tr
44	117	6.1	1230	2 T22458	hypothetical prote
45	117	6.1	1790	2 S67593	transport protein

ALIGNMENTS

RESULT 1

A45377

transcription factor CREB-2 - human

N:Alternate names: TAXREB67

C:Species: Homo sapiens (man)

C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 05-Nov-1999

C:Accession: A45377; I56787

R:Karpinski, B.A.; Morle, G.D.; Huggenvik, J.; Uhler, M.D.; Leiden, J.M.

Proc. Natl. Acad. Sci. U.S.A. 89, 4820-4824, 1992

A:Title: Molecular cloning of human CREB-2: an ATF/CREB transcription factor that can

A:Reference number: A45377; MUID:92279218

A:Accession: A45377

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-351 <KAR>

A:Cross-references: GB:M86842; NID:g181040; PIDN:AAA52071.1; PID:g181041

A:Note: the sequence is not complete

R:Tsujiimoto, A.; Niyunoya, H.; Morita, T.; Sato, T.; Shomotohno, K.

J. Virol. 65, 1420-1426, 1991

A:Title: Isolation of cDNA for DNA binding proteins which specifically bind to TAX-re

A:Reference number: I56787; MUID:91140735

A:Accession: I56787

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-351 <RES>

A:Cross-references: GB:D90209; NID:g220087; PIDN:BAA14234.1; PID:g220088

C:Genetics:

A:Gene: GDB:CREB2; ATF2; CRE-BP1; TREB7

A:Cross-references: GDB:I28011; OMIM:123811

A:Map position: 2q32-2q32

C:Keywords: transcription factor

Query Match 13.7%; Score 364; DB 2; Length 351;
Best Local Similarity 28.7%; Pred. No. 1.3e-09;
Matches 117; Conservative 53; Mismatches 122; Indels 116; Gaps 22;

QY 19 EMPVVTQDQFGDLKS---TSRHGDESLSL-----OPOGAT-----L 53

Db 3 EMSLSLSVLGDLMSPPDPSPGLGAEISGLDDYLVAKHFKPHGFSDDKAKASSEWL 62

QY 54 KLEPFF-----EDVLGAEMW-ESSDLGSF-LDALGDNNHRLHPPFESNLLEFSLIT- 102

Db 63 AVDGLVSPNNSKEDAFSGTDWMLKMKDLKEFDLDA-----LLGIDDLFTM 108

QY 103 PDDSTVSKD-----ILSSTLQFPQTPQVINIPLVASHGAEDFSAETEFENHLLSPDS-- 152

Db 109 PDDLTLTLDLDCDLFAPLVQETNKPQPTVN-PI-----GHL--PESIT 149

QY 153 -PEQVAPVINLEPVELTASHMTVISPDLGLGGMELASLSLTFELDFVNFNDNSAVGSGTG 211

Db 150 KPDQVAPFTFLQPLPLSPGVLS-STPDHSF-SLELGSF-SLELGSF-VDITEGDRKPDYAYVAMI-- 204

```

QY 19 EMPVQT--DQFGDLKSTRRGGSLSLQDQ-----GATLKLEPEEDVLG-AEW----- 67
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 37 ELPEQRSGVGGGLDDPFDQSGGLGABESLGLDDYLEVAKHLKPHGFSSDKAGSEWPAMD 96
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 68 --MESSDGLSFLDALGDNHRLHPFESNLLFTSLITPDDSTVSKDILSSTL----- 117
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 GLASASTGK--EDAFSGTDWMLKMDLKEFDLFRMDDLETMPDELTTLTUDTCDLPA 155
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 118 -----QFTQPNYNIPLASHGAEDEFAETEFENHLSPPDS---PEQVAPVINLEP--- 164
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156 PLVQETNKEPPQTVN-PI-----CHL--PESLIKVDQVAPPTFLQPPPC 196
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 165 -----VEL-----TASHMTVISPDGLLGGMELASESITFTTEL 196
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197 SPGLVSLTPEHSFLELGEYDISEGDRKPDAAAYITLIPP-----CVKKE 242
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 197 DFNFNDSAVSGISGAELGLGSLVDVVESTISFGSPSPETSQSSIIESSPELYKVIS 256
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 243 DTPSDNS--GICMSPESYLGP-----QHSFSTSR-----APPD--NLPS 279
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 257 TSSIDASKRFSRYSRKSOKSVKTSDAKAPRKTTPAQVPVPEHVIMEHLDKKDRKKLQN 316
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 280 PGSGRSGPRPKPYD-----PPGVSLTAKVKTEKLDKLLKMEQN 318
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 317 KNAAIRYRMKKKGAGQIKGEEQLEELNTKLTKYVDLQREIKYMKNNLMEDVCRAKG 374
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 319 KTAATRYQRKRAQEQALTGTGECKLEKNEALKKEKADSLAKEIQYLKDLIEEVRKAR 376
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
D34223
transcription factor ATF-4 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 28-Aug-1998
C:Accession: D34223
R:Hai, T.; Liu, F.; Coukos, W.J.; Green, M.R.
Genes Dev. 3, 2083-2090, 1989
A:Title: Transcription factor ATF cDNA clones: an extensive family of leucine zipper
A:Reference number: A91622; MUID:90185187
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-68 <HA4>
C:Genetics:
A:Gene: GDB:ATF4; TYREB; TAXREB67
A:Cross-references: GDB:I32551
C:Keywords: DNA binding; transcription regulation

Query Match 8.58; Score 163; DB 2; Length 68;
Best Local Similarity 50.08; Pred. No. 0.0003;
Matches 33; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

QY 304 EHLDDKKDRKLLQNKAATRYRMKKKGAEQIKGEEQLEELNTKLTKYVDLQREIKYMK 363
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 EKLDKLKRMEQNKRAATRYQRKRAEQEALTGTGECKLEKNEALKKEKADSLAREKQYILK 62
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 364 NLMEDV 369
   ||: ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 63 DLLEEV 68

RESULT 4
T27532
hypothetical protein ZC376.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27532
R:McMurray, A.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z20384
A:Accession: T27532

```

A:Molecule type: DNA
A:Residues: 1-488 <MBL>
A:Cross-references: EMBL:Z77136; PIDN:CAR00883.1; GSPDB:GN00023; CESP:ZC376.7
A:Experimental source: clone ZC376
C:Genetics:
A:Gene: CESP:ZC376.7
A:Map position: 5
A:Introns: 32/3; 56/2; 181/3; 205/3; 232/3; 450/2

Query Match 7.7% Score 148.5; DB 2; Length 488;
Best Local Similarity 24.1%; Pred. No. 0.039;
Matches 91; Conservative 53; Mismatches 115; Indels 119; Gaps 21;

Qy 21 PVVDTGQ--FGDLKSTRHGGDESLQPQGATLKLEPFEDY----LGAEW-----M 68
 |: : | | | | | | | | | | | | | | | | :
Db 184 PLFQSPSKSAIDLQTGTSR---IDEYGMPQDR--KLKSFEMDIQEESKAVDWEAWNYL 238

Qy 69 ESSDGLGFLDALGDNRHLRHPFNLENLFSTLPDDSTVKSDILSSTLTQPTQPVN--- 125
 ||| : | | | | | | | | | | | | | | | | : :
Db 239 ESDD-----DVFKRPFAF---FKPEPMIMTSSDS-----LMTSTSPDGSGISLYD 281

Qy 126 --IPLASH-CAEDFSABTEFEN--HLSPDPSP---EQAVPINLEPVELTASHTVI 177
 ||| : | | | | | | | | | | | | | | | | :
Db 282 PMIPPPSHFFSNLSSSSSANLRLLSTPSAPMQOEHRAVRMHHDVDFSSGPLLCVP 341

Qy 178 DGLLGMELASESLTFELDVFNFNDNSAVGSIGGAEEULLGPSLVSYDDVESTIFSGPSSP 237
 : | | | | | | | | | | | | | | | | | | :
Db 342 -----NQEDVFD--DFIQRD-----DDEDYI----PAS- 365

Qy 238 ETSQSSIIESPELYKVISTSIDASKRFSPYSRSSKSQSVKTSDAKPRKTTPAQP- 296
 : | : | : | : | : | : | : | : | : | : | :
Db 366 -----EARTSSLRNKRSATPTYLRRRDSR---RSWTTPASDD 399

Qy 297 -VPEH-----VIMEHLDK--DRKKLOKNKAIRMYMKKGCEAGIKGEQEFLELN 345
 || : | : | | | | | | | | | | | | | | | :
Db 400 YFPEHQKFKRGVVLKPSVDDETDRRLMLNRIAAVRYREKRAEKGRKMFEFOEVADRN 459

Qy 346 TKLTKVVDDLQREIKYM 363
 | | | | | | | | | | | | | | | | |
Db 460 RILLQKERQLKREINSMK 477

RESULT 5
A37953
transcription regulator PAN-1 - golden hamster (fragment)
C:Species: Mesocricetus auratus (golden hamster)
C>Date: 06-Dec-1991 sequence_revision 06-Dec-1991 #text_change 02-Jul-1998
C:Accession: A37953
R:German, M.S.; Blanas, M.A.; Nelson, C.; Moss, L.G.; Rutter, W.J.
Mol. Endocrinol. 5, 292-299, 1991
A>Title: Two related helix-loop-helix proteins participate in separate cell-spe
A:Reference number: A37953; MUID:91246228
A:Accession: A37953
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-436 <GER>
C:Superfamily: human transcription factor 3

Query Match	7.5%	Score 145;	DB 2;	Length 436;
Best Local Similarity	23.1%;	Pred. No. 0.055;		
Matches	86;	Conservative	56;	Mismatches 171; Indels 60; Gaps 14;
Qy	27	QFGDLKSTR-----HGGD-----ESLSLP---QGATLKLEPPE--EDVLGAEMM	68	
Db	46	GTFGLOQQRMSYQLHGSEYNGTLPVSSFSAAPTGYGGASGHTPPVSGADSLMGTGRGT	105	
Qy	69	ESSDLGFLDAGNHRRLHPFENLNLEFTSLIITPDDSTVSKDILLSLTLPQPTQPVNIPL	128	
Db	106	TASSSG----DALGKALASISPDHSSNNFSP--SFSTPVSGPQGLPGTYSQMPRACAPSAL	160	
Qy	129	YASHGAEDFSAETEFENHLSPDPDSEQVAPVINLEPVELTASHMTVISPDP--GLIIGGMEL	186	

```

Db 161 SPTYDGLHGLQSMEDRLD-----EAIHVLRSHAVGTASDLHGLLPG--- 203
Qy 187 ASESLTTELDVFNFNDSAYGVSIGGA--BELLGSPLSVDDVESTISFSGCSPSPETTSQSSI 244
Db 204 -HGALTTSFPGPVLPGRGHAGLVGGGHPEDGLTSGTSLTHIASLPSQASSLPDJSQ---- 259
Qy 245 IESSPELYKVISTSSIDASKRRFSPYSRSSK-SKQSVKTSDA-----KAPRKTRTPAOP 296
Db 260 --RPPDSVGGIGRAGAPAGA--SEIKREKKDDEESTSVADAEEDKDKLAPRTRTSPDES 315
Qy 297 VPEHVIHEHLDDKDRKKLQNNKNAIRYRMKKKGAGGIGEEQELKNTKTKTKVDDLIQ 356
Db 316 TDEVLSLEEKDLRORERMANNARERVVRDINEAFRELGRICQLHLKSDKAQTKLLIIQ 375
Qy 357 REIKYMKNLMDV 369
Db 376 QAVQVILGLEQV 388

RESULT 6
T24446
hypothetical protein T04C10.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T24446
R:Burton, J.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z19891
A:Accession: T24446
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-208 <WTL>
A:Cross-references: EMBL:Z69895; PIDN:CAA93757.1; GSPDB:GN00028; CESP:T04C10
A:Experimental source: clone T04C10
C:Genetics:
A:Gene: CESP:T04C10.4
A:Map position: X
A:Introns: 91/3; 183/3

```

Query Match	7.4%	Score 143.5;	DB 2;	Length 208;
Best Local Similarity	27.0%;	Pred. No. 0.024;		
Matches	43;	Conservative 28;	Mismatches 69;	Indels 19; Gaps 3;
QY	230	SFSGPSSPETSSSTIESSPELYKVIWSSSI--DASKRFSYPSSRSSKSKSVKTSDAK--	285	
Db	43	SYFNPSYHSQSQHHILNSDVNFQMVPOSTSVPPDFCSIEPMETNVQAKEQILBEIVREC	102	
QY	286	--APRKTTPAQVPPEHVIMEH-----LDKKRKKLQNKNAAIRMKMKKGE	330	
Db	103	EEIERSSNSSASPASNWSDEHDSOSEKSYHPYKTPPEKKERKKAQNRLAATRYREKKRRE	162	
QY	331	AGQTKGEQEELNLTUKTKVDDQLQREIKYMKNNLMEDV	369	
Db	163	KEEAMTCLEGLSVNGKLGKQDVSELERIYRFFKKFMTM	201	

```

RESULT# 7
CAMP response element-binding protein ATF2 - rat
A39429
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A39429
R:Kageyama, R.; Sasai, Y.; Nakanishi, S.
J. Biol. Chem. 266, 15525-15531, 1991
A:Title: Molecular characterization of transcription factors that bind to t
A:Reference number: A39429; MUID:91332085
A:Accession: A39429
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-389 <RAG>
A:Cross-references: GB:M65148; NID:q206569; PTD:AAA42013.1; PID:q206570

```

C: Superfamily: CAMP response element-binding protein 1; fos/jun DNA-binding domain homolog
C: Keywords: DNA binding; nucleus; transcription regulation
F: 231-271/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 7.2%; Score 138.5; DB 1; Length 389;
Best Local Similarity 23.6%; Pred. No. 0.12;
Matches 77; Conservative 47; Mismatches 137; Indels 65; Gaps 14;

QY 82 DNERHLPFENLLEFSLTP--DDSTVSKDILSSFLQFTQTPVNIPLY---ASHGAE 136
DB 21 EDHLVHKHKKH---EMTLKFGPARNDSEVIVADQPTTRFLKNCSEVGLFNELASPFENE 77
QY 137 FSAETFEFENLSPDSEQVAPVINL---EPVELTASHMTVISPDGLLGGMELASESLTF 193
DB 78 FKASEDDIKMPLDLSPLATPIIRSKIEEPSVVEVTHQDSPLPH-----PESTT- 127
QY 194 TELDFVNFDSAVGSIAGAEELGSLVSDVSTISFSGPSSPETSSQSIIESSPELYK 253
DB 128 -----ND-----EKLVRPVTM--VPSVPGIPGSPSPQVQS---EAKMRLKA 164
QY 254 VIS-----TSSIDAKRFSYSSKSKQSVKTSDAKAPKRTTPAQP---VPE----- 299
DB 165 ALTQQHPVTVNGDTVKGHSGSLVRAOSEERPOSLOQPATSTTETPASPAHTTPQTQNTS 224
QY 300 -----HVIMEHLKDKRKKLO--NKNAAIRYRMKKKGAEQIGKEQELEELNTKTKVDD 354
DB 225 GRRRAANEDPDEKRRKFLERNRAASRCQKRWVQVSLKAEADLSLNGQLQSEVTL 284
QY 355 LQREIKYMKNLN--EDVCKAKGIQLK 378
DB 285 LRNEVAQLKQLLAHKDCPVTAMQKK 310

RESULT 8
C42026
cyclic AMP response element DNA-binding protein isoform 2 - mouse
C: Species: Mus musculus (house mouse)
C: Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999
C: Accession: C42026
R: Georgopoulos, K.; Morgan, B.A.; Moore, D.D.
Mol. Cell. Biol. 12, 747-757, 1992
A: Title: Functionally distinct isoforms of the CRE-BP DNA-binding protein mediate acti
A: Reference number: A42026; MUID: 92123199
A: Contents: EL4
A: Accession: C42026
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-358 <GEO>
A: Cross-references: GB:S76659; NID:G243430; PIDN:AAB21129.1; PID:G243431
A: Note: sequence extracted from NCBI backbone (NCBIN:76659, NCBI:P:76660)
C: Superfamily: CAMP response element-binding protein 1; fos/jun DNA-binding domain homolog
F: 199-239/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 7.1%; Score 137; DB 2; Length 358;
Best Local Similarity 23.2%; Pred. No. 0.13;
Matches 70; Conservative 45; Mismatches 127; Indels 60; Gaps 11;

QY 104 DDTVSKDILSSFLQFTQVNIPLY---ASHGAEFSAETEFENHLSPPDSEQVAPVI 160
DB 10 NDSVIVADQPTTRFLKNCSEVGLFNELASPFENEKASEDDIKMPLDLSPLATPII 69
QY 161 NL---EPVELTASHMTVISPDGLLGGMELASESLTFELDFVNFDSAVSIGGAELLG 217
DB 70 RSKIEEPSVVEVTHQDSPLPH-----PESTT-----SDSKLV 101
QY 218 SPLSVDDVESTISFSGPSSPETSSQSIIESSPELYKVIS-----TSSIDAKRFSYSR 271
DB 102 RPYTM--VPSVPGIPGSPSPQVQS---EAKMRLKALTOQHPPVTNGDIVKGHSGSLVR 156
QY 272 SSKSKQSVKTSDAKAPKRTTPAQP---VPE-----HVIMEHLKDKRKKLO--NKN 318
C: Accession: S05380; A34776; B34223

DB 157 TQSEESRQSLQOPATSTTETPASPAHTTPTQNTSGRRRAANEDPDEKRRKFLERNRA 216
QY 319 AATRYRMKKKGAEQIGKEQELEELNTKTKVDDIQRKIKYMKNLN--EDVCKAKGIQ 376
DB 217 AASRCRKRKRWVQVSLKAEADLSLNGQLQSEVTLRNEVAQLKQLLAHKDCPVTAMQ 276
QY 377 LK 378
DB 277 KK 278
RESULT 9
B42026
cyclic AMP response element DNA-binding protein isoform 1 - mouse
C: Species: Mus musculus (house mouse)
C: Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999
C: Accession: B42026
R: Georgopoulos, K.; Morgan, B.A.; Moore, D.D.
Mol. Cell. Biol. 12, 747-757, 1992
A: Title: Functionally distinct isoforms of the CRE-BP DNA-binding protein mediate act
A: Reference number: A42026; MUID: 92123199
A: Contents: EL4
A: Accession: B42026
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-456 <GEO>
A: Cross-references: GB:S76657; NID:G243428; PIDN:AAB21128.1; PID:G243429
A: Note: sequence extracted from NCBI backbone (NCBIN:76657, NCBI:P:76658)
C: Superfamily: CAMP response element-binding protein 1; fos/jun DNA-binding domain ho
F: 297-337/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 7.1%; Score 137; DB 2; Length 456;

Best Local Similarity 21.7%; Pred. No. 0.19;
Matches 85; Conservative 57; Mismatches 168; Indels 82; Gaps 15;

QY 52 TLKLEPFEED-VLGAEMWESSDLGSLF---DALGDNHERLHPFESNLLEFTSLITPDOST 107
DB 2 TLKFGPARNDSEVIVAD--QTPTTRFLKNCSEVGLFNELASPFEN---EFKASEDDIKK 56
QY 108 VSKDILSSFLQFTQVNIPLYASHGAEDFSAETEFEN----HLSPDSEQVAPVIN-- 161
DB 57 MPLDL-----SPLATPIIRSKIEEPSVVEVTHQDSPLPHPESTTSDEKEVLAQTA 107
QY 162 -----LEPVELTASHMTVISPDGLLGGMELASESLTFELDFVNFDSAVSIGGAEL 215
DB 108 QPISAIVRPASLOQPNVLLTSSDSSVLIQQAIVPSTSTVITQAPSSNRPIVPPGPFPL 167
QY 216 L-----GSPLSVDD-----VESTISFSGPSSPETSSQSIIES 247
DB 168 LLHLPLNGQTPVAIPASITSSNVHVPAAVPLVRPVTWVPSVPGIPGSPSPQVQS---EA 224
QY 248 SPELYKVIS-----TSSIDAKRFSYSSKSKQSVKTSDAKAPKRTTPAQP---VP 298
DB 225 KMRLKAALTQOHPPTVNGDTVKGHSGSLVRVTQSESRPQSLQQPATSTTETPASPAHTT 284
QY 299 E-----HVIMEHLKDKRKKLO--NKNAAIRYRMKKKGAEQIGKEQELEELNTKL 348
DB 285 QTQNTSGRRRAANEDPDEKRRKFLERNRAASRCRKRKRWVQVSLKAEADLSLNGQL 344
QY 349 KTKVDDIQRKIKYMKNLN--EDVCKAKGIQLK 378
DB 345 QSEVTLRNEVAQLKQLLAHKDCPVTAMQKK 376

RESULT 10
S05380
CAMP response element-binding protein 1 - human
N: Alternate names: transcription factor, TGACGCA-binding
N: Contents: CAMP response element-binding protein HB16
C: Species: Homo sapiens (man)
C: Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 29-Sep-1999
C: Accession: S05380; A34776; B34223

Search completed: August 8, 2000, 02:39:56
Job time: 11275 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2000, 02:33:46 ; Search time 35.03 Seconds
(without alignments)
335.280 Million cell updates/sec

Title: US-08-656-811A-1
Perfect score: 128
Sequence: 1 MELDLWSEDFQLAREWGLEM.....KYNKNLMEDVCKAGIQLKM 379

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_38:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	264	13.7	351	1 ATF4_HUMAN	P18848 homo sapien
2	257	13.3	349	1 ATF4_MOUSE	Q06507 mus musculus
3	142.5	7.4	374	1 FOS_TETFL	Q91496 tetradodon f
4	138.5	7.2	376	1 FOS_FUGRU	P53450 fugu rubrip
5	137	7.1	487	1 ATF2_CHICK	O93602 gallus galli
6	137	7.1	487	1 ATF2_HUMAN	P15336 homo sapien
7	137	7.1	487	1 ATF2_MOUSE	P16951 mus musculus
8	134.5	7.0	330	1 YDC3_SCHPO	Q10424 schizosacch
9	134	7.0	487	1 ATF2_RAT	Q00969 rattus norv
10	134	7.0	647	1 SKO1_YEAST	Q02100 saccharomyc
11	133	6.9	483	1 ATF4_HUMAN	P17544 homo sapien
12	127	6.6	181	1 ATF3_HUMAN	P18847 homo sapien
13	127	6.6	638	1 PAN1_RAT	P21676 rattus norv
14	124.5	6.5	2774	1 MAPA_RAT	P34926 rattus norv
15	121	6.3	721	1 ENP1_TORCA	P14400 torpedo cal
16	120.5	6.2	331	1 API_PIG	P56432 sus scrofa
17	120.5	6.2	654	1 TFE2_HUMAN	P15923 homo sapien
18	119.5	6.2	893	1 MYM2_CAEEL	P34531 caenorhabdi
19	119	6.2	439	1 MYC_HYLLA	P49033 hyllobates l
20	119	6.2	1528	1 TP2A_MOUSE	Q01320 mus musculus
21	118.5	6.1	181	1 ATF3_MOUSE	Q06765 mus musculus
22	118.5	6.1	1863	1 BRCL_HUMAN	P38398 homo sapien
23	118	6.1	551	1 YD23_YEAST	Q07657 saccharomyc
24	117.5	6.1	682	1 HSF4_HUMAN	O99081 homo sapien
25	117	6.1	1790	1 UO01_YEAST	P25386 saccharomyc
26	116.5	6.0	181	1 ATF3_RAT	P29596 rattus norv
27	116	6.0	334	1 API_RAT	P17325 rattus norv
28	116	6.0	3358	1 PGCV_MOUSE	Q62059 mus musculus
29	115.5	6.0	587	1 YID6_YEAST	P40535 saccharomyc
30	115	6.0	334	1 API_MOUSE	P05627 mus musculus
31	113.5	5.9	2025	1 TTC3_HUMAN	P53804 homo sapien
32	113.5	5.9	2805	1 MAPA_HUMAN	P78559 homo sapien
33	113	5.9	281	1 TFE1_MOUSE	P15806 mus musculus

RESULT 1

ID	ATF4_HUMAN	STANDARD;	PRT;	351 AA.
AC	P18848;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	CYCLOC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-4 (DNA-BINDING PROTEIN			
DE	TAXREB67) (CYCLIC AMP RESPONSE ELEMENT-BINDING PROTEIN 2) (CREB2).			
GN	ATF4.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 91140735.			
RA	Tsuimoto A., Niyunoya H., Morita T., Sato T., Shimotohno K.;			
RT	"Isolation of cDNAs for DNA-binding proteins which specifically bind			
RT	to a tax-responsive enhancer element in the long terminal repeat of			
RT	human T-cell leukemia virus type I.";			
RL	J. Virol. 65:1420-1426(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 92279218.			
RA	Karpinski B.A., Morle G.D., Huggenvik J., Uhler M.D., Leiden J.M.;			
RT	"Molecular cloning of human CREB-2: an ATF/CREB transcription factor			
RT	that can negatively regulate transcription from the cAMP response			
RT	element.";			
RN	Proc. Natl. Acad. Sci. U.S.A. 89:4820-4824(1992).			
RN	[3]			
RP	SEQUENCE OF 274-341 FROM N.A.			
RX	MEDLINE; 90185187.			
RA	Hal T., Liu F., Coukos W.J., Green M.R.;			
RT	"Transcription factor ATF cDNA clones: an extensive family of leucine			
RT	zipper proteins able to selectively form DNA-binding heterodimers.";			
RN	Genes Dev. 3:2083-2090(1989).			
RN	[4]			
RP	ERRATUM.			
RA	Hal T., Liu F., Coukos W.J., Green M.R.;			
RL	Genes Dev. 4:682-682(1990).			
CC	-1- FUNCTION: THIS PROTEIN BINDS TO A TAX-RESPONSIVE ENHANCER			
CC	ELEMENT IN THE LONG TERMINAL REPEAT OF HUMAN T-CELL LEUKEMIA			
CC	VIRUS TYPE I.			
CC	-1- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE)			
CC	(CONSENSUS: 5'GTGAGCT(A/G)(A/G)-3'), A SEQUENCE PRESENT IN MANY			
CC	VIRAL AND CELLULAR PROMOTERS.			
CC	-1- SUBUNIT: BINDS DNA AS A DIMER.			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.			
CC	-1- SIMILARITY: TO OTHER BZIP PROTEINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			

ALIGNMENTS

34	112.5	5.8	1065	1	SED4_YEAST
35	112	5.8	439	1	MYC_PANTR
36	112	5.8	507	1	YG46_YEAST
37	111.5	5.8	1616	1	P200_MYCGE
38	111.5	5.8	3685	1	DMD_HUMAN
39	111	5.8	211	1	CES2_CAEEL
40	111	5.8	2464	1	MAPB_MOUSE
41	110.5	5.7	326	1	FRA2_MOUSE
42	110.5	5.7	331	1	API_HUMAN
43	110.5	5.7	1938	1	MYSA_MOUSE
44	110.5	5.7	1938	1	MYSA_RAT
45	110	5.7	482	1	YSR2_CAEEL

P25365	saccharomyc
P23583	pan troglod
P53301	saccharomyc
Q49429	mycoplasma
P11532	homo sapien
Q94126	caenorhabdi
P14873	mus musculus
P47930	mus musculus
P05412	homo sapien
Q02566	mus musculus
P02563	rattus norv
Q09950	caenorhabdi

```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; D90209; BAA14234.1; -.
DR EMBL; M6842; AAA5071.1; -.
DR PIR; D34223; D34223.
DR HSP; P03069; IDGC.
DR TRANSFAC; T01303; -.
DR MIM; 604064; -.
DR PFAM; PF00170; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DNA_BIND 280 300
FT DOMAIN 306 334
FT CONFLICT 284 284 K -> R (IN REF. 2).
FT CONFLICT 290 290 T -> R (IN REF. 3).
FT CONFLICT 329 331 KEI -> REK (IN REF. 3).
FT CONFLICT 338 338 I -> L (IN REF. 3).
SQ SEQUENCE 351 AA; 38558 MW; 7A708C5CCD6ED7F8 CRC64;

Query Match 13.7%; Score 264; DB 1; Length 351;
Best Local Similarity 28.7%; Pred. No. 3.3e-09;
Matches 117; Conservative 53; Mismatches 122; Indels 116; Gaps 22;

QY 19 EMPVVTQDGFGLDKS---TSRHGGDESLSL-----QPGAT-----L 53
DB 3 EMSFLSEVLVGLDMLSPFDPGLGAESLGLDDYLEVAKHFKPHGFSDDKAGSSEWL 62
QY 54 KLEFF-----EDVLGAEMW-EESDLGSF-LDAFGDNHERLHPFESNLEFSLIT- 102
DB 63 AVDGLVSPNSNKEDAFSGTDMLEKMDLKEFDLDA-----LLGIDDLLETM 108
QY 103 PDDTSVKD-----ILSTLFQTPQVNIPLYASHGAEDFAETEFENHLSPPDS-- 152
DB 109 PDDLTLTDDTCDFAPLVQETNKQPQTUN-PI-----GHL--PESLT 149
QY 153 -PEQVAPVINLEPVELTASHMTVISPDLGLGGMELASESLTFTLDFVNFNDSAVGSI 211
DB 150 KPDQVAPFTFLQPLSPGVLS-STPDHSF-SLELGSE-VDITEGDRKPDYATVAMI-- 204
QY 212 AEELGSLPVDVDESISFGSPSPETSQSSIIETSPELYKVIISTSIDASKRFSYSR 271
DB 205 -----PQCIKEED-----TPSDNDSGIC-MSPESYL-----GSPQHSSTR 239
QY 272 SSKSKQSVKYSDAKAPKTRTPAQVPVEHVM-----EHLDDKDKKLNKNAAIRYRMK 326
DB 240 GSPNR-SLSPGVLCGSRPKPYDPGCKVAAKVGKELDKLKKMEQNKTAATRIQK 298
QY 327 KKGAAQIKGEEQEELENTKLTQVDDLQREIKYMKNLMDYCKAKG 374
DB 299 KRAQEALTGCKEKEKNEALKERADSLAKEIQYLKDLIEVRKARG 346

RESULT 2
ID ATFA_MOUSE STANDARD; PRT; 349 AA.
AC Q06507; Q61906;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-4 (C/EBP-RELATED ATF)
DE (C/ATF) (TAXREB67 HOMOLOG).
GN ATF4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ADIPOCYTE;
RX MEDLINE; 93281642.
RA Valjejo M., Ron D., Miller C.P., Habener J.F.;
RT "c/ATF, a member of the activating transcription factor family of
RT DNA-binding proteins, dimerizes with CAAT/enhancer-binding proteins

```

```

RT and directs their binding to cAMP response elements.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:4679-4683(1993).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-CD-1; TISSUE-EMBRYO;
RX MEDLINE; 92335183.
RA Chevray P.M., Nathans D.;
RT "Protein interaction cloning in yeast: identification of mammalian
RT proteins that react with the leucine zipper of Jun.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:5789-5793(1992).
CC -1- FUNCTION: BINDS TO ASYMMETRIC CAMP RESPONSE ELEMENTS (CRE) AS
CC A HETERODIMER AND TO PALINDROMIC CRE'S AS A HOMODIMER.
CC -1- SUBUNIT: BINDS DNA AS A HOMO- AND HETERODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: BRAIN, HEART, LIVER, SPLEEN, LUNG, AND MUSCLE.
CC -1- SIMILARITY: TO OTHER bZIP PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL; L13791; AAA40476.1; -.
CC EMBL; M94087; AAA53043.1; ALT_INIT.
CC HSP; P03069; IDGC.
CC MGD; MGI:88096; ATP4.
CC PFAM; PF00170; bZIP; 1.
CC PROSITE; PS00036; bZIP_BASIC; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DNA_BIND 278 298
FT DOMAIN 304 332
FT CONFLICT 345 345 LEUCINE-ZIPPER (PROBABLE).
FT CONFLICT 349 349 Q -> K (IN REF. 2).
SQ SEQUENCE 349 AA; 38355 MW; 0C3F89574051C7B9 CRC64;

Query Match 13.3%; Score 257; DB 1; Length 349;
Best Local Similarity 27.4%; Pred. No. 8.6e-09;
Matches 115; Conservative 40; Mismatches 123; Indels 142; Gaps 18;

QY 19 EMPVVTQDGFGLDKS---TSRHGGDESLSLQPO---GATLKLEPFEEDVLG-AEW--- 67
DB 3 EMSFLSEVLVGLDMLSPFDPGLGAESLGLDDYLEVAKHFKPHGFSDDKAGSSEWPAM 62
QY 68 -----MESSDLGSFLDALGDNHERLHPFESNLEFSLITPDDSTVSKDILSSTL----- 117
DB 63 DDGLASASDTGK-EDAFSGTDMLEKMDLKEFDLAFRMDDLTMDPELTLTDDTCDL 121
QY 118 -----OPPTOPVNIPLYASHGAEDFAETEFENHLSPPDS---PEQVAPVINLEP- 164
DB 122 FAPLVQETNKPEPQTUN-PI-----GHL--PESLIKVDQVAPFTFLQPF 162
QY 165 -----VEL-----TASHTVISPDLGLGGMELASESLTFT 194
DB 163 PCSPGVLSSTPEHSFSLSELGSEVDISEGDRKPDSAAYITLIPP-----CVK 208
QY 195 ELDFVNFNDSAVSIGAEELGSLVSDVESTISFGSPSPETSQSSIIETSPELYKV 254
DB 209 EEDTPSDNDS--GTCMSPESYLGSP-----QHSPTSR-----APPD--NL 245
QY 255 ISTSSIDASKRFSYRSKSKQSVKTSDAKAPKTRTPAQVPVEHVMELDKDKKRL 314
DB 246 PSPGSGSGSPRPYD-----PPGVSLTAKVKTETKDKLKKME 284
QY 315 QNKNAAIRYRMKKGAAQIKGEEQEELENTKLTQVDDLQREIKYMKNLMDYCKAKG 374
DB 285 QNKTAATRIYRQKRAEQEALTGCKEKEKNEALKERADSLAKEIQYLKDLIEVRKARG 344

RESULT 3
FOS_TETFL

```

```

ID FOS_TETFL STANDARD; PRT; 374 AA.
AC 091496;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE P55-C-FOS PROTO-ONCOGENE PROTEIN (CELLULAR ONCOGENE C-FOS).
GN FOS
OS Tetraodon fluviatilis (Puffer fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
OC Percomorpha; Tetraodontiformes; Tetraodontidae;
OC Tetraodon.
RN [1]
RP SEQUENCE FROM N.A.
RA Chang M.S., Chang G.D., Huang F.L., Huang C.J., Lo T.B.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: NUCLEAR PHOSPHOPROTEIN WHICH FORMS A TIGHT BUT NON-
CC COVALENTLY LINKED COMPLEX WITH THE C-JUN/AP-1 TRANSCRIPTION
CC FACTOR. C-FOS HAS A CRITICAL FUNCTION IN REGULATING THE
CC DEVELOPMENT OF CELLS DESTINED TO FORM AND MAINTAIN THE SKELETON.
CC IT IS THOUGHT TO HAVE AN IMPORTANT ROLE IN SIGNAL TRANSDUCTION,
CC CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC !- SUBUNIT: HETERODIMER (BY SIMILARITY).
CC !- SUBCELLULAR LOCATION: NUCLEAR.
CC !- SIMILARITY: BELONGS TO THE BZIP FAMILY. FOS SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U53520; AAB07359.1; -
DR HSSP; P01100; LFOS.
DR PFAM; PF00170; BZIP; 1
DR PRINTS; PR00042; LEUZIPPROS.
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW Proto-oncogene; Nuclear protein; Phosphorylation; DNA-binding.
FT DNA_BIND 122 143 BASIC MOTIF.
FT DOMAIN 148 176 LEUCINE-ZIPPER.
SQ SEQUENCE 374 AA; 40525 MW; 4DE2CB426D9FEEB9 CRC64;

Query Match 7.4%; Score 142.5; DB 1; Length 374;
Best Local Similarity 26.6%; Pred. No. 0.065;
Matches 45; Conservative 35; Mismatches 76; Indels 13; Gaps 4;

Qy 208 SIGGAELLGSPVDDVESTISFGSPSPETSOSIIESSPELYKVIS--TSSIDASKR 265
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 28 SPAGSYSSMGSPQSDLTDLTAS----SASFVPTVTAISTSPDLQWVPLVSSVAPSR 83
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 266 FSPYSRSKSKOSVKTSDAKAPKTRTPAQVPVHMEHLDDKDRKLNKNAIRYRM 325
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 84 AHPTSPSPYKRTVMRSASPHAKRGV----EQTPEEEKKRIRERNKQAAKCRN 139
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 326 KKGGAQGIKEOELELNTKLTKVDDLORE----IKYMKNLMDVCK 371
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 RRRLTDSLQAEITDLEAKSSLDNDIANLLKERELEFILAHHQPICK 188
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
FOS_FUGRU
ID FOS_FUGRU STANDARD; PRT; 376 AA.
AC P53450;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE P55-C-FOS PROTO-ONCOGENE PROTEIN (CELLULAR ONCOGENE C-FOS).
GN FOS
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
OC Percomorpha; Tetraodontiformes; Tetraodontidae;
OC Takifugu.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96202283.
RA Trower M.K., Orton S.M., Purvis I.J., Sanseau P., Riley J.,
RA Christodoulou C., Burt D., See C.G., Elgar G., Sherrington R.,
RA Rogaev E.I., St George-Hyslop P.H., Brenner S., Dykes C.W.;
RT "Conservation of synteny between the genome of the pufferfish (Fugu
RT rubripes) and the region on human chromosome 14 (14q24.3) associated
RT with familial Alzheimer disease (AD3 locus).";
RL Proc. Natl. Acad. Sci. U.S.A. 93:1366-1369(1996).
CC !- FUNCTION: NUCLEAR PHOSPHOPROTEIN WHICH FORMS A TIGHT BUT NON-
CC COVALENTLY LINKED COMPLEX WITH THE C-JUN/AP-1 TRANSCRIPTION
CC FACTOR. C-FOS HAS A CRITICAL FUNCTION IN REGULATING THE
CC DEVELOPMENT OF CELLS DESTINED TO FORM AND MAINTAIN THE SKELETON.
CC IT IS THOUGHT TO HAVE AN IMPORTANT ROLE IN SIGNAL TRANSDUCTION,
CC CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC !- SUBUNIT: HETERODIMER (BY SIMILARITY).
CC !- SUBCELLULAR LOCATION: NUCLEAR.
CC !- SIMILARITY: BELONGS TO THE BZIP FAMILY. FOS SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U40757; AAC59778.1; -
DR HSSP; P01100; LFOS.
DR TRANSFAC; T02205; -
DR PFAM; PF00170; BZIP; 1
DR PRINTS; PR00042; LEUZIPPROS.
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW Proto-oncogene; Nuclear protein; Phosphorylation; DNA-binding.
FT DNA_BIND 123 144 BASIC MOTIF.
FT DOMAIN 149 177 LEUCINE-ZIPPER.
SQ SEQUENCE 376 AA; 40826 MW; BFC28534431DB491 CRC64;

Query Match 7.2%; Score 138.5; DB 1; Length 376;
Best Local Similarity 27.0%; Pred. No. 0.11;
Matches 53; Conservative 36; Mismatches 76; Indels 31; Gaps 7;

Qy 198 FVNFN-----DSAVG-----SIGGAELLGSPVDDVESTISFGSPSPETSQ 241
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 FTSFNAECDSSSRCSPVGNLYYPSPAGSYSSMGSPQSDFTDLTAS----SASFPT 58
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 242 SSIIESSPELYKVIS--TSSIDASKRFPYSRSKSKOSVKTSDA-KAPKTRTPAQVP 298
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 VTAISTSPDLQWVPLVSSVAPSHRAHPYSPSPYKRTVMRSASPHAKRGSRVQPTP 118
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 299 EHVIMEHLDDKDRKLNKNAIRYMKKGEAQGIKEOELELNTKLTKVDDLORE 358
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 EEE-----EKKRIRERNKQAAKCRNRRLTDTLQAEITDLEAKSSLDNDIANLLKE 173
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 359 ---IKYMKNLMDVCK 371
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 KERLEFILAHHQPICK 189

RESULT 5
ATF2_CHICK
ID ATF2_CHICK STANDARD; PRT; 487 AA.
AC O93602;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-2 (ACTIVATING
DE TRANSCRIPTION FACTOR 2).

```

GN ATF2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huglier S., Baguet J., Perez S., van Dam H., Castellazzi M.;
 RT "Transcription factor ATF2 cooperates with vJun to promote growth-
 factor independent proliferation in vitro and tumor formation in
 vivo";
 RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RL CC -1- FUNCTION: A TRANSCRIPTIONAL ACTIVATOR, PROBABLY CONSTITUTIVE,
 WHICH BINDS TO THE CAMP RESPONSE REGION (CRE) (CONSENSUS:
 5'GTGACGT(A/C)(A/G)-3'); A SEQUENCE PRESENT IN MANY VIRAL AND
 CC CELLULAR PROMOTERS (BY SIMILARITY).
 CC -1- SUBUNIT: BINDS DNA AS A DIMER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. ATF SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Y17724; CAA76838.1; -;
 DR HSP; P08047; ISP2.
 DR PFAM; PF00170; bZIP; 1.
 DR PROSITE; PS00096; zf-C2H2; 1.
 DR PROSITE; PS00036; BZIP_BASIC; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
 KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
 KW Zinc-finger; Metal-binding.
 FT ZN_FING 7 31 C2H2-TYPE.
 FT DOMAIN 333 356 BASIC MOTIF.
 FT DOMAIN 362 390 LEUCINE-ZIPPER.
 FT SEQUENCE 487 AA; 52406 MW; 41F42734D9C6A146 CRC64;
 SQ
 Query Match 7.18; Score 137; DB 1; Length 487;
 Best Local Similarity 21.38; Pred. No. 0.19;
 Matches 84; Conservative 60; Mismatches 163; Indels 88; Gaps 15;
 QY 52 TLKLEPFEED-VLCAEWESDLGSFL--DALGDNHERLHPFESNLLEFTSLTPDDST 107
 DB 34 TLKLEPARNDSVIVAD--QTPTTRFLKNCVEVGLFNLASPFEN---EFKASEDDIKK 88
 QY 108 VSKDILSSTLQFPQPVPNIPLYASHGAEDFSAETEFENHLSPP-SDSPEQVAPVINLE--- 163
 DB 89 MPLDL-----SPLATPIIRNKIEPSVVEVTHQDSPLPHPESTINDEKESLQTA 139
 QY 164 -----PVELTASHMTVISPDLGLGMEELASESLFTTDLDFVNFDSAVGSIIGGAEL 215
 DB 140 QPTSTIVRPASLPVNPVLLTSSDSSVITQAIPTSTSTVITQATPSSNRPIVPPGPFPL 199
 QY 216 L-----GSLPSVDD-----VESTISFGSPSPETSQS----- 242
 DB 200 LLHLPLNGQTMVPAIPASITNSVHVPAAPLVLRVPTVTPSPGIPGSPSPQVQSEAKLR 259
 QY 243 ---SIIESSPELYKVIYSTSIDASK-----RFSYSPRSKSKQSVKTSDAKAPKTRT 292
 DB 260 LKAALTQHQVQV-----TNGDTAKHPSGLVTRTQSEPRPSLQOQATSTTETPASPQA 313
 QY 293 PAQVPP-----EHVIMEHLDDKKRKLQ-NKNAAIYRMYKKGAGQIGEEQELN 345
 DB 314 PTQQTPTNGRRRAANEDPDKKRKFELNRRAAASRCQKRKVVWVSGLEKKAEDLSLN 373
 QY 346 TLKTKVDLDLQREIKYMKNLW--EDVCCKAGIOLK 378
 DB 374 GOLONEVTLRLNRVAQLKLLLAHKQCPVTAMQKK 408

RESULT 6
 ATF2_HUMAN
 ID ATF2_HUMAN STANDARD; PRT; 487 AA.
 AC P15336;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-2 (ACTIVATING
 DE TRANSCRIPTION FACTOR 2) (CAMP RESPONSE ELEMENT BINDING PROTEIN CRE-
 DE BPL1) (HB16).
 DE BPL1 (HB16).
 GN ATF2 OR CREB2 OR CREBPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Maekawa T., Sakura H., Kanei-Ishii C., Sudo T., Yoshimura T.,
 RT Fujisawa J.I., Yoshida M., Ishii S.;
 RT "Leucine zipper structure of the protein CRE-BPL binding to the
 RT cyclic AMP response element in brain";
 RL EMBO J. 8:2023-2028(1989).
 CC [2]
 CC SEQUENCE OF 193-487 FROM N.A.
 CC MEDLINE; 90705810.
 CC "A cDNA for a human cyclic AMP response element-binding protein which
 CC is distinct from CREB and expressed preferentially in brain";
 CC Mol. Cell. Biol. 10:1347-1357(1990).
 CC -1- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE)
 CC VIRAL AND CELLULAR PROMOTERS.
 CC -1- SUBUNIT: BINDS DNA AS A DIMER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: MXBP AND CRE-BPL MAY BE MEMBERS OF A FAMILY
 CC OF MXBP/CRE-BP PROTEINS GENERATED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: ABUNDANT EXPRESSION SEEN IN THE BRAIN.
 CC -1- PTM: PHOSPHORYLATION OF THR-69 AND THR-71 BY MAPK14 CAUSES
 CC INCREASED TRANSCRIPTIONAL ACTIVITY. ALSO PHOSPHORYLATED AND
 CC ACTIVATED BY JNK.
 CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. ATF SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X15875; CAA33886.1; -;
 DR EMBL; M31630; AAA35951.1; -;
 DR PIR; S05380; S05380.
 DR HSP; P08047; ISP2.
 DR TRANSFAC; T00167; -;
 DR MIM; 123811; -;
 DR PFAM; PF00170; bZIP; 1.
 DR PFAM; PF00096; zf-C2H2; 1.
 DR PROSITE; PS00036; BZIP_BASIC; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
 KW Transcription regulation; DNA-binding; Activator; Phosphorylation;
 KW Nuclear protein; Alternative splicing; Zinc-finger; Metal-binding.
 FT ZN_FING 7 31 C2H2-TYPE.
 FT DOMAIN 333 356 BASIC MOTIF.
 FT DOMAIN 362 390 LEUCINE-ZIPPER.
 FT MOD_RES 51 51 PHOSPHORYLATION (BY MAPK14).
 FT MOD_RES 53 53 PHOSPHORYLATION (BY MAPK14).
 FT SEQUENCE 487 AA; 52249 MW; BEE4D17FE8DB7CD7 CRC64;


```
ID YDC3_SCHPO STANDARD; PRT; 330 AA.
AC Q10424;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HPOTHEICAL 36.2 KDA PROTEIN C25G10.03 IN CHROMOSOME I.
GS SPAC25G10.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z70691; CAA94632.1; -.
CC HSP; P03059; 2DGC.
CC PROSITE; PS00036; BZIP_BASIC; 1.
KW Hypothetical protein; Transcription regulation; DNA-binding;
FT DNA_BIND 270 288 BASIC MOTIF (BY SIMILARITY).
FT SEQUENCE -330 AA; 36189 MW; A681434C779DF960 CRC64;
SQ
Query Match 7.0%; Score 134.5; DB 1; Length 330;
Best Local Similarity 22.0%; Pred. No. 0.16;
Matches 72; Conservative 54; Mismatches 130; Indels 71; Gaps 10;
QY 95 LEFTSLITPDSTVSKDILSTLQFPQPVNIPL- - - - -YASHGAEDFSAETEFENHLSPP 150
DB 14 LKFDVDPVSD- - -FSKDDLAQLNFTNPNFLDLEPSMLSEGYGFVQSGSSNSKKQ 71
QY 151 D- - - - -SPEQAPVINLEPVELTASHMTVISPDGLGGMELASESLTFTELD- - - - -FVN 200
DB 72 EKNVQOQNPKEIKISTLQVKEEVNTESAPLNATGNFSSANPASIDLAYLDLKLTLPLD 131
QY 201 FNDASVGSIGGAELLGSLV- - - - -DDVESTISFGPSS- - - - - 236
DB 132 HSKETQKTSQRELFQKSSVASASKDNVSSSILQGSASSKLLPDQSARQHOVLVGQT 191
QY 237 - - - - -PTSQSSIIESPVELYK- - - - -IST-SSIDAKRFSPYSRSSKSKQS- - - - -V 279
DB 192 AIPTEASSSSINNTPLQAPVSSFDQAFNPNLSTFPADPLASVSSPSLSYKGAQSPNA 251
QY 280 KTSDAKAPKTRTPAQVPPEHVIHEHLKDKKLLKLNKAAIRYMRKKGAEQKIGEEQ 339
DB 252 NSKRTKATSAIRTAEE- - - - -EDKRRRTAASAREFRICKLKEQLERTAK 297
QY 340 ELEELNTKLTQVDDLOREIKYMKNL 366
DB 298 ELTEKVAILETRVRELEMENNWLKGLI 324
```

RESULT 9

```
ID ATF2_RAT STANDARD; PRT; 487 AA.
AC Q00969; Q62870;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-2 (ACTIVATING
DE TRANSCRIPTION FACTOR 2) (CAMP RESPONSE ELEMENT BINDING PROTEIN CRE-
```

```
DE BP1).
GN ATF2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC SYRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;
RA Muramatsu S.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE-BRAIN;
RA Kagevama R., Sasai Y., Nakanishi S.;
RL "Molecular characterization of transcription factors that bind to the
RT CAMP responsive region of the substance P precursor gene. CDNA
RT cloning of a novel C/EBP-related factor."
CC J. Biol. Chem. 266:15525-15531(1991).
CC -1- FUNCTION: A TRANSCRIPTIONAL ACTIVATOR, PROBABLY CONSTITUTIVE,
CC WHICH BINDS TO THE CAMP RESPONSIVE REGION (CRE) (CONSENSUS:
CC 5'GTGACGT(A/C)(A/G)-3'); A SEQUENCE PRESENT IN MANY VIRAL AND
CC CELLULAR PROMOTERS.
CC -1- SUBUNIT: BINDS DNA AS A DIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: ISOFORM 1 (SHOWN HERE) AND
CC ISOFORM 2: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. ATF SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U38938; AAA93263.1; -.
CC EMBL; M65148; AAA42013.1; -.
CC PIR; A39429; A39429.
CC HSP; P08047; ISP2.
CC TRANSFAC; T01382; -.
CC PFAM; PF00170; BZIP; 1.
CC PFAM; PF00096; ZF-C2H2; 1.
CC PROSITE; PS00036; BZIP_BASIC; 1.
CC PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
KW Transcription regulation; DNA-binding; Activator; Phosphorylation;
KW Nuclear protein; Alternative splicing; Zinc-finger; Metal-binding.
FT ZN_FING 7 31
FT DNA_BIND 333 356 BASIC MOTIF.
FT DOMAIN 362 390 LEUCINE-ZIPPER.
FT MOD_RES 51 51 PHOSPHORYLATION (BY MAPK14) (BY
FT MOD_RES 53 53 SIMILARITY).
FT MOD_RES 53 53 PHOSPHORYLATION (BY MAPK14) (BY
FT VARSPLIC 132 229 MISSING (IN ISOFORM 2).
FT SEQUENCE 487 AA; 52286 MW; 4ED95B106DF5F9EE CRC64;
SQ
Query Match 7.0%; Score 134; DB 1; Length 487;
Best Local Similarity 21.9%; Pred. No. 0.29;
Matches 86; Conservative 58; Mismatches 166; Indels 82; Gaps 16;
QY 52 TLKLEPFED-VLGAEMWESSDLGSLF- - -DALGDNHERLHPESNLLETSLITPDST 107
DB 34 TLKFGPARNDSVIVAD- - -QTPTPTFLKNCSEEVGLFNELASPFEN- - -EFKASEDDIKK 88
QY 108 VSKDILSSTLQFPQPVNIPLVASHGAEDFSAETEFENHLSPP- - -PEQVAPVIN- - 161
DB 89 MPLDL- - - - -SPLATIIIRSKTEEPSVVTTHQDPLPHESTYNDKEIPLAQTA 139
QY 162 - - - - -LEPVELTASHMTVISPDGLGGMELASESLTFTELDVFNFNDSAVSIGGAEL 215
```

```
Db 140 QPTSAIVRPASLQVNPVLLTSSDSSVILIIQAVPSPSTSTVITQAPSSNRPIVVPGRPL 199
QY 216 L-----GSLPVSDD-----VESTISFGSPSPETSQSSIIES 247
Db 200 LHLPLNGQTMPIAIPATISSTNVHVPAAVPLVRPVTMPVSPGIPGSPSPQVS---EA 256
QY 248 SPELYKVIS-----TSSIDASKRSPYSRSKSKSVKTSDAKAPRTRTPAQP---VP 298
Db 257 KURLAALTOQHPPTVNGDTYKVGHSGLVRAQSESRQSLQQPATSTETPASPAHTTP 316
QY 299 E-----HVIMEHLDKDKRKLLQ--NKNAIRYRMKKKGGAQKIGEEQEBLEELNKL 348
Db 317 QFQNTSGRRRAANEDPDEKRKFELERNRAASRCQRKRYVWVQSLKKAEDLSSINGOL 376
QY 349 KTKVDLDLQREIKYMKNL--EDVCAKAGIQLK 378
Db 377 QSEVTLRNEVAQLKOLLAHKDCPVTAMQKK 408

RESULT 10
SKOL_YEAST STANDARD; PRT; 647 AA.
AC Q02100;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CRE-BINDING BZIP PROTEIN SKO1.
OS SKO1 OR ACR1 OR YNL167C OR NI702.
GC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-W303-1A;
RX MEDLINE; 93065191.
RA Nehlin J.O., Carlberg M., Ronne H.;
RT "yeast SKO1 gene encodes a bZIP protein that binds to the CRE motif
and acts as a repressor of transcription.";
RL Nucleic Acids Res. 20:5271-5278(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93078739.
RA Vincent A.C., Struhl K.;
RT "ACR1, a yeast ATF/CREB repressor.";
RL Mol. Cell. Biol. 12:5394-5405(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE; 96287653.
RA Nasr F., Becam A.-M., Herbert C.J.;
RT "The sequence of 36.8 kb from the left arm of chromosome XIV reveals
24 complete open reading frames: 18 correspond to new genes, one of
which encodes a protein similar to the human myotonic dystrophy
kinase.";
RL Yeast 12:169-175(1996).
CC -!- FUNCTION: BINDS TO THE CRE MOTIF 5'-TGACGTCA-3' AND ACTS AS A
REPRESSOR OF TRANSCRIPTION OF THE SUC2 GENE AND MOST PROBABLY
OTHER GENES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: TO OTHER BZIP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67875; CAA48074.1; -
DR EMBL; S49588; AAB24288.1; -
DR EMBL; X92317; CAA63272.1; -
DR EMBL; Z71443; CAA96054.1; -
```

```
DR PIR; S26386; S26386.
DR PIR; A45028; A45028.
DR HSSP; P05412; 1FOS.
DR TRANSFAC; T01306; -.
DR SGD; L0001909; SKO1.
DR PEAR; PF00170; bZIP; 1.
DR PROSITE; PS00036; BZIP_BASIC; FALSE_NEG.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
FT DNA_BIND 431 456 BASIC MOTIF.
FT DOMAIN 457 478 LEUCINE-ZIPPER.
SQ SEQUENCE 647 AA; 70192 MW; 3E0B8C72A6CE14B CRC64;

Query Match 7.0%; Score 134; DB 1; Length 647;
Best Local Similarity 21.1%; Pred. No. 0.43;
Matches 75; Conservative 60; Mismatches 132; Indels 88; Gaps 14;

QY 45 SLQOGATLKLEPFEEDVLGAEMESSDLGSLDALGDHNLHPDESLLFTSLITPD 104
Db 169 SSSPGTSLIRPNSV-----TTSNG---NFPTNDSQMPGFLLLSK--SGLTPN 217
QY 105 DSTVSKDILSSTLQFPTQPVNIPLYASHGAEDFSAETEFENHLSPDPSPEQ-----VAP 158
Db 218 ESNIRTLGTPGIL---TQSYNYPVLPSINKNITGSKNVKSVTVNGSTENHPHNIMP 274
QY 159 VINLPVELTASHMVTVISPDGLLGWELASELSLTTELDVFNPNFNSAVSGISGAELLGS 218
Db 275 TVNGTPLTGLSLLNLPSTGVLANPVFKSTPTT-----NTTIDT----- 314
QY 219 PLSVDVESTISFGSPSPETSQSSIIE--SPELYKVIS-----TSSIDASKR 265
Db 315 -----VNNSISNSN-FSPNTSTKAAVKMDNPAETFAHNAIHSAHNHKENENLTQIENDDO 367
QY 266 FSPYSRSKSKQSVKTSDAKAPRK-----RTPAQ-----PVPEHVIMEHLDK 308
Db 368 FNNKTRKRRMSSTSSTSKSKNSISRKNSAVTAPAKDDVENKKSNNVTLDENEE 427
QY 309 KDRKKLQ-----NKNAIRYRMKKKGGAQKIGEBEELNKLTKYVDLQREI 359
Db 428 QERKRKEFLERNVAASKRKRKKEYIKKIENDLQFYE-----SEYDULTQVI 475

RESULT 11
ATFA_HUMAN STANDARD; PRT; 483 AA.
AC P17544;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE TRANSCRIPTION FACTOR ATF-A AND ATF-A-DELTA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90301459.
RA Gaire M., Chatton B., Kedinger C.;
RT "Isolation and characterization of two novel, closely related ATF
cDNA clones from HeLa cells.";
RL Nucleic Acids Res. 18:3467-3473(1990).
CC -!- FUNCTION: THIS PROTEIN BINDS THE CAMP RESPONSE ELEMENT (CRE)
VIRAL AND CELLULAR PROMOTERS.
CC -!- SUBUNIT: BINDS DNA AS A DIMER.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- ALTERNATIVE PRODUCTS: ATF-A AND ATF-A-DELTA, WHICH ARE IDENTICAL,
EXCEPT FOR ADDITIONAL 21 AMINO ACIDS IN ATF-A (RESIDUES 114-134)
MAY BE DERIVED FROM A COMMON PRIMARY TRANSCRIPT BY ALTERNATIVE
SPLICING EVENTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
```


ID AC PAN1_RAT STANDARD; PRT; 638 AA.
 DT P21676;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE TRANSCRIPTIONAL REGULATORY PROTEIN PAN-1 (FRAGMENT).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE; 90346284.
 RA Nelson C., Shen L.-P., Meister A., Fodor E., Rutter W.J.;
 RT "Pan: a transcriptional regulator that binds chymotrypsin, insulin,
 and AP-4 enhancer motifs.";
 RL Genes Dev. 4:1035-1043(1990).
 CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT BINDS TO THE CONSENSUS
 SEQUENCE CAC/GCTGT/C PRESENT, IN THE CHYMOTRYPSIN, INSULIN, AP-4,
 AND SEVERAL OTHER GENE ENHANCER MOTIFS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: PAN-1 AND PAN-2 MAY BE GENERATED BY
 ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 TRANSCRIPTION FACTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X62323; CA44199.1; .
 DR PIR; B35816; B35816.
 DR HSP; P10085; 1MDY.
 DR TRANSFAC; T00674; .
 DR PFAM; PF00010; HLH; 1.
 DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
 DR Transcription regulation; DNA-binding; Nuclear protein; Repeat;
 KW Alternative splicing.
 FT NON_TER 1 1
 FT REPEAT 180 184 S-S-S-G-D.
 FT REPEAT 314 318 S-S-S-G-D.
 FT DOMAIN 377 412 LEUCINE-ZIPPER (POTENTIAL).
 FT DNA_BIND 533 587 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 SQ SEQUENCE 638 AA; 66350 MW; 74489490C97D2B5 CRC64;

Query Match 6.6%; Score 127; DB 1; Length 638;
 Best Local Similarity 21.7%; Pred. No. 1.1;
 Matches 96; Conservative 54; Mismatches 146; Indels 146; Gaps 23;

QY 4 DLWSEDFQ-----LARENGLEMPVQV-----TDGQFGDLKSTSR-----HGGD----- 41
 DB 219 ELWPPQGVGFGMLGGDSAPLAPAGSSSVSSGAFGLQGDQRMGVQLHGSVEVNGTLPA 278
 QY 42 -ESLSLOP--QGATLKLEPPE--EDVLGAEMWESSDLGFLDALGDNHRLHPFESNLL 95
 DB 279 VSSFSAPGTYSGTSGHTPPVSGADSLGTRGTASSG---DALGK----- 322
 QY 96 EFTSLTPDDSTVKDILSSTLQFPPTQVPIPLVASHGAEDFSAETEPENHLSPPDSPEQ 155
 DB 323 ALASIYSPD-----HSSNNFSPSPS-----TPVGSF-Q 349
 QY 156 VAPVNLPEVELTASHMTVISP--DGLLGGMELASESLTFTLDFVFNDSAV-----GS 208
 DB 350 GLPGTSQMP-----RAGAPSLSPNDAGLHGLSKMEDRL-----DEAHLVLRSHA 395
 QY 209 IGGAEELGSLPSVDDVESTISFSGP-----SSPE---TSQSII----- 245
 DB 396 VGTASELHG--LLPGHSLTLTTSFAGPSLGGRHAGLVSGSHPDGLTSGASILLHNHSLP 453

QY 246 -----BSSPELYKVISSIDASKRSPSPSRSSKSKQSVKT-----SDAKAP 287
 DB 454 SPPSLPDLSSQRPDPDSFSGLRAGVTAGA--SEIKREEKEDEEVTSTVADAEEDKDKLVP 511
 QY 288 RKTRTPAQVPPEHVIMEHLDDKDKKLNKNAAIYRMKKKGGAQGIKGEEOELEELNTK 347
 DB 512 R-TRTSS--TDEVLSLEEKDLDRERRMANNARERVRVDINEAFRELGRMQLHLKSDK 568
 QY 348 LKTKVDDLQREIKYMKNLMDV 369
 DB 569 AQTKLLILQQAQVQVILGLEQOV 590

RESULT 14
 MAPA_RAT
 ID MAPA_RAT STANDARD; PRT; 2774 AA.
 AC P34926;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2].
 GN MAP1A.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE; 92355629.
 RA Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
 RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
 one messenger RNA.";
 RL J. Biol. Chem. 267:16561-16566(1992).
 CC -1- FUNCTION: STRUCTURAL PROTEIN INVOLVED IN THE FILAMENTOUS
 CROSS-BRIDGING BETWEEN MICROTUBULES AND OTHER SKELETAL ELEMENTS.
 CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN-ASSOCIATE
 WITH MAP1A AND MAP1B PROTEINS.
 CC -1- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
 APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
 THEIR MORPHOLOGY.
 CC -1- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
 FOR THE BINDING OF MAP1A TO MICROTUBULES.
 CC -1- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
 CC -1- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
 FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
 BOTH MAP1A AND MAP1B.
 CC -1- SIMILARITY: TO MAP1B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M83196; AAB48069.1; .
 DR PIR; A43359; A43359.
 KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN 72465 2774 MAP1 LIGHT CHAIN LC2.
 FT DOMAIN 309 496 LYS-RICH (BASIC).
 FT REPEAT 336 541 11 X 3 AA REPEATS OF K-K-[DE].
 FT REPEAT 336 338 1.
 FT REPEAT 415 417 2.
 FT REPEAT 420 422 3.
 FT REPEAT 424 426 4.
 FT REPEAT 427 429 5.
 FT REPEAT 431 433 6.
 FT REPEAT 436 438 7.
 FT REPEAT 440 442 8.
 FT REPEAT 444 446 9.
 FT REPEAT 449 451 10.

```
FT REPEAT 539 541 11.
SQ SEQUENCE 2774 AA; 299526 MW; 3DEF74427BA9D7D7 CRC64;

Query Match 6.58; Score 124.5; DB 1; Length 2774;
Best Local Similarity 22.3%; Pred. No. 11;
Matches 91; Conservative 50; Mismatches 142; Indels 125; Gaps 20;

QY 30 GDLKSTSRHG-GDESLSLQO-----GATLLEPEEDVLGAENMES-SDLGSP 76
Db 1045 GPEGTSEKGGPPRSPQADMPVSTAGGTCTIQLLPQDKAIVFETGEAGNLGAG 1104
QY 77 L---DALGDHNERLHPESNLLEFT-SLITPDD-----STVSKDILSLFLQ--FPTQ 123
Db 1105 TLPCEVSTSEATEPEKQDEVLRITDQSLSPEDAESLSVSVSPD---TTKQEAATPRSP 1161
QY 124 VNI-----PLYA-----SHGAEDFAETEE---NHLSPDP-----S 152
Db 1162 CSLKEQPHKDLMPWSPEDTQSLSFSESPSKETSLDISKQLSPESLGTQFGELNLG 1221
QY 153 PEQVAPVI-----NLEPVLTASHMTVISPDCLLGGMELASESLTFELDFVNFDSA 205
Db 1222 KEERGPNVKAEDDSCHLAPVSIPIPHRATVSP-----STDETP 1259
QY 206 VGSIGGAELGLSPSLVDVESTISFSGSPSPETSSQSIIESSPELYKVISSIDASKR 265
Db 1260 AGTLPGG-SFSHSALSVDKRHSPEITGCGHFMTSDSLTKSP----- 1303
QY 266 FSPYRSKSKQSVKTS-DAKAPKRTTP-----AQVPVHEVIMEHLD-----KKD 310
Db 1304 ----SLSSPAMEDLAVEWECKAPGKEREPELKSRTQKQGLPEKVAVVEQDLIIHQKD 1359
QY 311 RKLQNKAAIRYRMKKKGGAQIGKEQELEELNKLTKTVDDLORE 358
Db 1360 -GALDEENKFGROQDKTPKQKGRDLDEKDTAAELDKGPEPEKDLDR 1406

RESULT 15
ENPL_TORCA STANDARD; PRT; 721 AA.
AC PI4400; 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE ELECTROMOTOR NEURON-ASSOCIATED PROTEIN 1 (FRAGMENT).
OS Torpedo californica (Pacific electric ray).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Neoselachii; Squalia; Hypnosqualea; Pristioraja;
OC Batoidea; Torpediniformes; Torpedinidae; Torpedo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90091742.
RA Ngsee J.K., Scheller R.H.;
RT "Isolation and characterization of two homologous cDNA clones from
RT Torpedo electrotomotor neurons.";
RL DNA 8:555-561(1989).
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH MEMBRANES OF INTRACELLULAR
CC ORGANELLES.
CC -1- MISCELLANEOUS: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE
CC SEQUENCE KKE, REPEATED BUT NOT AT FIXED INTERVALS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; M30270; AAA49279.1; --
DR PIR; A33319; A33319.
DR HSSP; P11387; 1A36.
KW Neurone; Repeat.
```

```
FT NON_TER 1 1
FT SIMILAR 233 582
FT 80% IDENTITY TO ELECTROMOTOR NEURON-
FT ASSOCIATED PROTEIN 2.
FT GLU/LYS-RICH (BASIC).
FT 11 X 3 AA APPROXIMATE REPEATS.
FT 1.
FT 2.
FT 3.
FT 4.
FT 5.
FT 6.
FT 7.
FT 8.
FT 9.
FT 10.
FT 11.
FT NON_TER 721 721
SQ SEQUENCE 721 AA; 81241 MW; B53E69A9F309CE00 CRC64;

Query Match 6.3%; Score 121; DB 1; Length 721;
Best Local Similarity 20.5%; Pred. No. 2.9;
Matches 88; Conservative 68; Mismatches 157; Indels 116; Gaps 19;

QY 38 HGGDESL-----LQPOGATLKLPEEEDVLGAENME---SSDLGSLDALGDHNERLHP 89
Db 316 HNGADNLPGVNSLLQRLKIAELEEEQSQSTANSWMKNLISPELCVFLNVPENVNLQ 375
QY 90 -----FESNLLEFTSL-----ITPDD--STVSKDILSLSTLQFPQPV-NIPLYASHGAE 135
Db 376 NFRVRRNTEETCTLQVNLKLCVKPELFTVVG-NVIDPVLFOKMGVGRLEMYILNSVK 434
QY 136 DFSAETEFENHLSPPDSPPEQVAPVINLEPVLTASHMTVIS----- 176
Db 435 GSKELQFFMQHWSNNKAKTGLILPNGKDAEISPPYLTSSLSLWHPANPSEKIVRALF 494
QY 177 ----PD-GLLGMELASESLTFTELDFVNFNDSAVGSGIGGAELGLSPS----- 221
Db 495 PGNAPQVNLIDGLE-----KLKHLDFLKH-----PVVTQKELIASAAPTQVQAKLKQ 542
QY 222 -VDDVESTISFGSPSPETSSQSIIESSPELYKVISSIDASKRFPSPYSSKSKOSVK 280
Db 543 WSDSKESLKSNSRPSVGKGVKDKVKEETPELTPTAVSHQEAUNE-KPQKVEKKEKPVVK 601
QY 281 TSDAKAPKRTTTPAQVPPE-----HVTMEHLDKDKKKLQNKNAAIYRMKKK 328
Db 602 -----KERPRTELQSKPEKDAKAKAQEELKMQDKELKSKPKPLKEKIVKK 655
QY 329 GEAQIGKEQELEELNKL-----KTKVDDLOREIKYMKNL 365
Db 656 -EVKAKPPEKKKEKDKVESAKPDRKEKAVIKKEKVPKEKPKKEVKDKV--KDV 712
QY 366 MEDVCKARG 374
Db 713 KKEV-KKKG 720

Search completed: August 8, 2000, 06:12:24
Job time: 13118 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2000, 01:43:29 ; Search time 39.91 Seconds
(without alignments)
658.422 Million cell updates/sec

Title: US-08-656-811a-1
Perfect score: 1928
Sequence: 1 MELDLWSEDFQLAREWGLEM.....KYMKNLMEDVCVAKGIQLKMK 379

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 59334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_12:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1923	99.7	378	5	Q16946
2	248	12.9	381	11	Q61328
3	246.5	12.8	354	13	Q9W610
4	188.5	9.8	182	4	Q9Y201
5	162	8.4	84	11	Q70191
6	148.5	7.7	488	5	Q23272
7	143.5	7.4	208	5	Q22156
8	137.5	7.1	494	4	Q13814
9	137	7.1	358	11	Q64091
10	137	7.1	456	11	Q64090
11	137	7.1	487	13	Q93602
12	135	7.0	486	13	Q91576
13	134	7.0	487	11	Q62870
14	133	6.9	448	11	Q64089
15	131	6.8	2390	11	Q08592
16	129.5	6.7	457	5	Q19521
17	127.5	6.6	1414	11	Q63330
18	127	6.6	460	5	Q17801
19	126.5	6.6	338	5	Q16949

20	126	6.5	699	11	Q35451	mus musculus
21	126	6.5	938	4	Q13111	homo sapien
22	125	6.5	507	11	Q92125	mus musculus
23	124.5	6.5	1142	4	O60732	homo sapien
24	124.5	6.5	1142	4	O75451	homo sapien
25	123	6.4	329	4	Q13000	homo sapien
26	123	6.4	613	5	Q9XZS8	drosophila
27	121	6.3	424	10	Q41109	phaseolus v
28	120.5	6.2	780	10	O04346	arabidopsis
29	119.5	6.2	319	13	Q91639	xenopus lae
30	119	6.2	341	3	Q12191	saccharomyc
31	119	6.2	700	4	Q99941	homo sapien
32	119	6.2	700	4	Q13289	homo sapien
33	118.5	6.1	181	11	O62281	mus musculus
34	118.5	6.1	590	3	Q08490	saccharomyc
35	118.5	6.1	1151	5	O9XUR2	caenorhabdi
36	118	6.1	378	10	O23964	heliobacter
37	118	6.1	547	3	Q9Y7Y1	schizosacch
38	117.5	6.1	427	10	Q40587	nicotiana t
39	117.5	6.1	601	5	O76379	echinococcu
40	117.5	6.1	656	5	Q9Y114	drosophila
41	117.5	6.1	1604	4	O95643	homo sapien
42	117.5	6.1	1792	4	O15129	homo sapien
43	117	6.1	355	3	P78962	schizosacch
44	117	6.1	703	4	Q14343	homo sapien
45	117	6.1	1230	5	O20626	caenorhabdi

ALIGNMENTS

RESULT 1

Q16946 ID Q16946 PRELIMINARY; PRU; 378 AA.
AC Q16946;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DE APCREB2.
GN APCREB2.
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidea;
OC Aplysiidae; Aplysia.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-CENTRAL NERVOUS SYSTEM;
RX MEDLINE; 96107336
RA BARTSCH D., GHIRARDI M., SKEHEL P.A., KARL K.A., HERDER S.P., CHEN M.,
RA BAILEY C.H., KANDEL E.R.;
RT "Aplysia CREB2 represses long-term facilitation: relief of repression
RT converts transient facilitation into long-term functional and
RT structural change."
RL Cell 83:979-992(1995).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; U40851; AAA92437.1; -
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; BZIP; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 378 AA; 41692 MW; A359586E CRC32;

Query Match 99.7%; Score 1923; DB 5; Length 378;
Best Local Similarity 100.0%; Pred. No. 1.1e-123;
Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELDLWSEDFQLAREWGLEMPVVDGQGDLSKTSRHGGDESLSLQPGATLKLEPFE 60
|||||

Db 1 MELDLWSEDFQLAREWGLEMPVVDGQGDLSKTSRHGGDESLSLQPGATLKLEPFE 60
|||||

Qy 61 DVLGAENWESSDLSGFLDALGDNHERLHPFESNLLEFTSLITPDDSTVSKDLSLTLOFP 120
|||||

Db 61 DVLGAENWESSDLSGFLDALGDNHERLHPFESNLLEFTSLITPDDSTVSKDLSLTLOFP 120
|||||

QY	121	TOPNIPLYASHGAEDFSAETEFENHLSPDPSPEQVAPVINLEPVELTASHMTVISPDL	180
Db	121	TOPNIPLYASHGAEDFSAETEFENHLSPDPSPEQVAPVINLEPVELTASHMTVISPDL	180
QY	181	LGGMELASESLTTETLDFVNFNDSAVGSI GGAEELGSPLSVDDVESTISFSGPSSPETS	240
Db	181	LGGMELASESLTTETLDFVNFNDSAVGSI GGAEELGSPLSVDDVESTISFSGPSSPETS	240
QY	241	QSSIIESSPELYKVIWTSIDASKRFSYRSKSKSQSVKTSDAKAPRTRTPAQVPVEH	300
Db	241	QSSIIESSPELYKVIWTSIDASKRFSYRSKSKSQSVKTSDAKAPRTRTPAQVPVEH	300
QY	301	VIMEHLDDKRRKKLQNKNAAIRYMKKKKGAGGKGEQELBELNKKTKVDDDLQREIK	360
Db	301	VIMEHLDDKRRKKLQNKNAAIRYMKKKKGAGGKGEQELBELNKKTKVDDDLQREIK	360
QY	361	YMKNLAMEDVCKAKGIOLK	378
Db	361	YMKNLAMEDVCKAKGIOLK	378

RESULT 2

Q61328	PRELIMINARY;	PRT;	381 AA.
ID	Q61328		
AC	Q61328;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)		
DE	ACTIVATING TRANSCRIPTION FACTOR 4 (ACTIVATING TRANSCRIPTION FACTOR 4 (ATF4)).		
DE	ATF4 OR ATF-4.		
GN	Mus musculus (Mouse).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN-129/SV;		
RC	MEDLINE; 92066493.		
RT	MELNICKI L.M., PRUITT S.C.;		
RT	"Isolation and nucleotide sequence of a murine CDNA homologous to		
RT	human activating transcription factor 4.";		
RL	Nucleic Acids Res. 19:6332-6332(1991).		
RN	[2]		
RP	SEQUENCE OF 119-381 FROM N.A.		
RA	NISHIZAWA M., NAGATA S.;		
RL	Submitted (MAR-1998) to the EMBL/GenBank/DDAJ databases.		
RN	[3]		
RP	SEQUENCE OF 119-381 FROM N.A.		
RX	MEDLINE; 92183900.		
RA	NISHIZAWA M., NAGATA S.;		
RT	*CDNA clones encoding leucine-zipper proteins which interact with G-		
RT	CSF gene promoter element 1-binding protein.";		
RL	FEBS Lett. 299:36-38(1992).		
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).		
CC	-1- SIMILARITY: TO OTHER BZIP PROTEINS.		
DR	EMBL; X61507; CAA43723.1; -.		
DR	EMBL; AB012277; BAA25314.1; -.		
DR	MGD; MGI:88096; Atf4		
DR	PROSITE; PS00036; BZIP_BASIC; 1.		
DR	PFAM; PF00170; bzip; 1.		
DR	CDNA-binding; Nuclear protein.		
SW	SEQUENCE 381 AA; 41810 MW.		
QY			C8136FE7 CRC32;

Qy	68	--MESSDGLSFLDALGDNHRLHPFFESNLLLEFTSLTTPDDSTVSKDILSSTL-----	111
Db	97	GLASASDTGK--EDAFSGTDWMLKMDLKEFDLALFRMDLFTMPDELITTLDDTCDLFA	155
Qy	118	-----QFTQPQWNIPLYASHGAEDFSAETEFENILSPDPS---PEQVAPVINLEP---	164
Db	156	PLVQETNKEPPQTVN--PI-----GHL--PESLIKVDQVAPFTFLQPPFC	196
Qy	165	-----VEL-----TASHMTVISPOGLLGGMELASESLTFTTEL	196
Db	197	SPGVLSSTPEHSFSELGSEVDISEGDRKPDSAAYITLIPP-----CVKEE	242
Qy	197	DFVFNDSAVCSIGCAEELGLSPSLVDVVESTISFGSPSPETSQSIIESSPELYKKVIS	256
Db	243	DTPSDNDS--GICMSPESYLGSP-----QHSPTSFR-----APPD--NLPS	279
Qy	257	TSSDASKRFPSPYSRSKSKQSVKTSDAKAPRKTRTPAQVPVPHVMEHLDKDRKKLQGN	316
Db	280	PGSGRSGPRKPYD-----PGVSLTAKVKYTEKLDKKLKKMEGN	318
Qy	317	KNAAIRYRMKKKGGAQOIKGEQELEBELNTKTKYVDDDLQREIKYKMNLMEDVCRAKG	374
Db	319	KTAATRYRQKRAEQEALGETCKELEKKNKALKEKADSLAKEIQYLKDLIEVRKARG	376

RESULT 3

Q9W610

PRELIMINARY;

PRT; 354 AA.

AC Q9W610;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)

DE ATF4.

DE ATF4.

GN ATF4.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=EYE;

KA MURATA T., SUZUKI H., OGINO H., YASUDA K.;

RT "Isolation and analysis of embryonic expression pattern of chicken ATF4."

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.

DR EMBL; AB013138; BAA76466.1; -.

DR PROSITE; PS00036; BZIP_BASIC; 1.

DR DNA-binding; Nuclear protein.

KW SEQUENCE 354 AA; 38990 MW; 6202DDAA0 CRC32;

Query Match 12.9%; Score 248; DB 11; Length 381;
Best Local Similarity 27.0%; Pred. No. 1.5e-09;
Matches 113; Conservative 40; Mismatches 125; Indels 140; Gaps 18;


```
Db 232 -----QISPNLSGPNP-----FPTDACSRSKPYDHPAEKVASAK 272
QY 281 TSDAKAPKTRTPAQPVPHEVIMEHDKDRKLNKNAAIYRMRKKKGAEQIGKEE 340
Db 273 VKGEK-----KIDKKLKEQNKTAATRYQRKRAEQEALSSECRD 313
QY 341 LEELNTKLTQVDDLOREIKYMKNLMDVCKAK 373
Db 314 LBQKNQALKERADSLSKETQYKLDLIEEVKAK 346

RESULT 4
QY2D1 PRELIMINARY; PRT; 182 AA.
AC QY2D1;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE LEUCINE-ZIPPER PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA KOHROKI J., TANAKA K.;
RT "cDNA clone encoding leucine-zipper protein."
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; AB021663; BAA78477.1; -
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 182 AA; 20428 MW; 61A81591 CRC32;

Query Match 9.8%; Score 188.5; DB 4; Length 182;
Best Local Similarity 30.8%; Pred. No. 6.4e-06;
Matches 45; Conservative 30; Mismatches 56; Indels 15; Gaps 2;

QY 234 PSSPTSSSISSPELYKVISSIDASKRFSYRSKSKQSVKTSDAKAPKTRTP 293
Db 39 FSLPLSLFSLPQPVL-----DTLDLLAICRNEAQEEVGMPPLPPQPPPP 89

QY 294 AQPVE-----HVIMEHDKDRKLNKNAAIYRMRKKKGAEQIGKEE 347
Db 90 SPQPQSLAPYPHPATTRGDRKQKRDQNKSAALYRQRKRAEGEAGECEQLEARNRE 149

QY 348 LKTKVDDLOREIKYMKNLMDVCKAK 373
Db 150 LKRAESVEREIQYVKDLLIEVYKAR 175

RESULT 5
O70191 PRELIMINARY; PRT; 84 AA.
AC O70191;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE ATPA (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA NISHIZAWA M., NAGATA S.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92183900.
RA NISHIZAWA M., NAGATA S.;
RT "cDNA clones encoding leucine-zipper proteins which interact with g-
```

```
RT CSF gene promoter element 1-binding protein."
RL FEBS Lett. 299:36-38(1992).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; AB012276; BAA25313.1; -
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; BZIP; 1.
KW DNA-binding; Nuclear protein.
FT NON_TER 1
SQ SEQUENCE 84 AA; 9893 MW; 04BE945C CRC32;

Query Match 8.4%; Score 162; DB 11; Length 84;
Best Local Similarity 43.3%; Pred. No. 0.00015;
Matches 29; Conservative 22; Mismatches 16; Indels 0; Gaps 0;

QY 307 DKDKRKLQNKNAAIYRMRKKKGAEQIGKEEQLNLTNLTQVDDLOREIKYMKNL 366
Db 11 DRKQKRDQNKSAALYRQRKRAEGEAGECEQGLEARNRELRRAESVEREIQYVKDLL 70

QY 367 EDVCKAK 373
Db 71 IEVYKAR 77

RESULT 6
Q23272 PRELIMINARY; PRT; 488 AA.
AC Q23272;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TRENBLrel. 09, Last annotation update)
DE ZC376.7 PROTEIN.
GN ZC376.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
OC Rhabdita; Rhabditidae; Rhabditinae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans."
RL Nature 368:32-38(1994).
DR EMBL; Z77136; CAB00883.1; -
SQ SEQUENCE 488 AA; 56019 MW; 80F3CB5D CRC32;

Query Match 7.7%; Score 148.5; DB 5; Length 488;
Best Local Similarity 24.1%; Pred. No. 0.012;
Matches 91; Conservative 53; Mismatches 115; Indels 119; Gaps 21;

QY 21 PVVQTGGQ--FGDLKSTSRHGDESLSLQPGQATLKLEPFEEVDY-----LGAEN-----M 68
Db 184 PLFQSPSKSAIDQLATGTSR---IDEXGMPPQDR--KLKSFENDIEQESKAVDWEANNHYL 238

QY 69 ESSDLGSLFDALGDNHERLHPESNLLLEFTSLTPDDSTVSKDILSSTLQFTQPNV--- 125
Db 239 ESDD-----DVFKRPEAF---FKPEPMINTSSDS-----LMTSSTSPSDGSLYD 281
```

```

230 QY SPFGSPSPETSQSIITESSPELYKVTSTSSI--DASKRPSVRSRKSQSKSVKTSDAK--285
43 Db SYFNPSXSHSQOQHLLNSVNFQWPMQPSVDPDFCSIEPMETVQAKELIEIVREC 102
286 QY --APRTRTPAQPVPEHVMIEH-----LDKDRKKLONKNAAIRYMKKKRGE 330
103 Db EETERASNSASPAWSNDSDHDSQSEKSYHPVKTPEKKERKKAONRLAATRYRKKRR 162

```

[illegible]

DE ACTIVATING TRANSCRIPTION FACTOR 2
GN (CYCLIC AMP RESPONSE ELEMENT DNA-BINDING PROTEIN ISOFORM 2).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
RN Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RP [1]
RX SEQUENCE FROM N.A.
RA MEDLINE; 92123199.
RA GEORGOPOULOS K., MORGAN B.A., MOORE D.D.;
RT "Functionally distinct isoforms of the CRE-BP DNA-binding protein
RL mediate activity of a T-cell-specific enhancer.";
RL Mol. Cell. Biol. 12:747-757(1992).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; S76659; AAB21129.1; -.
DR HSSP; P03069; 2DGC.
DR MGD; MGI:109349; Atf2.
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; bZIP; 1.
DR DNA-binding; Nuclear protein.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 358 AA; 38978 MW; D4B31477 CRC32;

Query Match 7.1%; Score 137; DB 11; Length 358;
Best Local Similarity 23.2%; Pred. No. 0.05; Mismatches 45; Indels 60; Gaps 11;
Matches 70; Conservative 45; Mismatches 45; Indels 60; Gaps 11;

QY 104 DDSTVSKDILSTLQFPQPNIPLY---ASHGAEDFSAETEFENHLSPPDPSPQEVAPVI 160
DB 10 NDSVIVADQTPTRFLKNCVEVGLFELASPFENEFKASEDDIKKPLDLSPLATPII 69
QY 161 NL---EPVELTASHMTVISPDGLGGMELASELTFTELDFVNFNDSAVSIGGAELLG 217
DB 70 RSKTEEPSVETTHQDSPLPH-----PESTT-----SDEKLV 101
QY 218 SPLSVDDVESITSPGSPSPSTOSSIIESPELYKVIS-----TSSIDASKRFPSPYR 271
DB 102 RPVTM--VPSVPGIIPGSPSPQVQS---EAKMLKAAALTOQHPVTVNGDTVKGHSGILVR 156
QY 272 SSKSKQSVKTSDAKAPKTRTPAQP---VPE-----HVIMEHLDKDKKKLQ-NKN 318
DB 157 TQSESRPQSLQQPATSTETPASPATHTPQTQNTSGRRRAANEDDPDEKRRKFLERNR 216
QY 319 AAIRYRMKKGEAGQIKGEQEELNKLTKTKYDDDLQREIKYMKNL--EDVCCKAKGIQ 376
DB 217 AASRCQRKRYVQVQSLKKAEDLSNLQLOSEVTLRNEVAQLKQLLAHKDCPVTAMQ 276
QY 377 LK 378
DB 277 KK 278

RESULT 10
Q64090 PRELIMINARY; PRT; 456 AA.
AC Q64090;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE ACTIVATING TRANSCRIPTION FACTOR 2
DE (CYCLIC AMP RESPONSE ELEMENT DNA-BINDING PROTEIN ISOFORM 1).
GN ATf2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE; 92123199.
RA GEORGOPOULOS K., MORGAN B.A., MOORE D.D.;
RT "Functionally distinct isoforms of the CRE-BP DNA-binding protein
RL mediate activity of a T-cell-specific enhancer.";
RL Mol. Cell. Biol. 12:747-757(1992).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; S76657; AAB21128.1; -.
DR HSSP; P03069; 2DGC.
DR MGD; MGI:109349; Atf2.
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; bZIP; 1.
DR DNA-binding; Nuclear protein.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 456 AA; 48931 MW; 8C79E121 CRC32;

Query Match 7.1%; Score 137; DB 11; Length 456;
Best Local Similarity 21.7%; Pred. No. 0.068; Mismatches 57; Indels 82; Gaps 15;
Matches 85; Conservative 57; Mismatches 168; Indels 82; Gaps 15;

QY 52 TLKLEPFEE-VLGAEMMESDLGSFL---DALGONHERLHFPESNLLEFTSLITPDDST 107
DB 2 TLKFGPARNDSVIVAD--QTPTPTFLKNCVEVGLFELASPFEN--EFKASEDDIKK 56
QY 108 VSKDILSTLQFPQPNIPLYASHGAEDFSAETEFEN-----HLSPPDPSPQEVAPVIN-- 161
DB 57 MPLDL-----SPLATPIIRSKIEEPSVVETTHQDSPLPHPESTTSDKEVPLAQT 107
QY 162 -----LEPVELTASHMTVISPDGLGGMELASELITELDFVNFNDSAVSIGGAEL 215
DB 108 QPSTAIVPASLQVNPVLLTSSDSSVITQOAVPSPTSTVITQAPSSNRPIVPVPGPPL 167
QY 216 L-----GSPLSVDD-----VESTISFGSPSPSTQSSIIES 247
DB 168 LLLHPNGQTPMVAIPASITSSNVHVPAAVPLVPVTVPSVPGIIPGSPSPQVQS---EA 224
QY 248 SPELYKVIS-----TSSIDASKRFPSPYRSKSKQSVKTSDAKAPKTRTPAQP---VP 298
DB 225 KMLKAAALTOQHPVTVNGDTVKGHSGILVRTQSESRPQSLQQPATSTETPASPATHTP 284
QY 299 E-----HVIMEHLDKDKKKLQ-NKNAIRYRMKKKGAEQIKGEQEELNKL 348
DB 285 QTQNTSGRRRAANEDDPDEKRRKFLERNRAASRCQRKRYVQVQSLKKAEDLSNLQ 344
QY 349 KTKVDDLQREIKYMKNL--EDVCCKAKGIQLK 378
DB 345 QSEVTLRNEVAQLKQLLAHKDCPVTAMQK 376

RESULT 11
O93602 PRELIMINARY; PRT; 487 AA.
ID O93602;
AC O93602;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE ACTIVATING TRANSCRIPTION FACTOR 2.
GN ATf2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RX SEQUENCE FROM N.A.
RA HUGUIER S., BAGUET J., PEREZ S., VAN DAM H., CASTELLAZZI M.;
RT "Transcription factor ATF2 cooperates with vjun to promote growth-
RT factor independent proliferation in vitro and tumor formation in
RT vivo.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; Y17724; CAA76838.1; -.
DR HSSP; P08047; ISP2.
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; bZIP; 1.
DR PFAM; PF00096; zf-C2H2; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 487 AA; 52406 MW; 52FB7330 CRC32;

```

Query Match          7.1%; Score 137; DB 13; Length 487;
Best Local Similarity 21.3%; Pred. No. 0.074;
Matches 84; Conservative 60; Mismatches 163; Indels 88; Gaps 15;

QY 52 TLKLEPFEED-VLGAEMWESSDLGSFL---DALGDNHERLHPFESNLEFTSLITPDST 107
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 34 TLKFGPARNDSVIVAD--QTPTPTFRFKNCEEVGLFNLASPFEN---EFKASEDDIKK 88
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 108 VSKDILSTLQFPQPNIPLYASHGAEDFSAETEFENHLSP-PDSPEQVAPVINLE--- 163
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 89 MPLDL-----SPLATPIIRKIEEPSVWETHQDSPLPHESTNDKEVSLQOTA 139
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 164 -----PVELTASHMTVISPDLGLGMELASESLFTTDLDFVNFDSAVGSIIGGAEL 215
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 140 QPTSTIVRPASLPVNPVLLTSSDSSVLIQQAIPSTSTVITQAPSSNRPIVPGFPPL 199
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 216 L-----GSPLSVDD-----VESTISFGSPSPSPETSQS----- 242
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 200 LLHLPNGQTMVPAIPASITNSNVHVPAAVPLVRPVTWVPSTIPGIPGSPSPQVQSEAKLR 259
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 243 ---SIIESSPELYKVISSIDASK-----RFSYRSKSKSQSVKTSDAKAPKTRT 292
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 260 LKAALTOQHPOV-----TNGDTAKGHPGSLVRTQSEPRPQSLQQAIPSTSTETPASPQ 313
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 293 PAQVP-----EHVIMEHLDDKDRKKLQ-KNNAAIRYRMKKKGAEQEELEELN 345
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 314 PTQQTPTNGRRRAANEDPDEKRRKFLERNRAAASRCQKRVVWSLEKKAEDLSLN 373
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 346 TLKTKVDDDLQREIKYMKNLK--EDVCCKAKGIQK 378
   | | : | | | | | : | | | : | | : | | : | | : | | : | | : | | :
Db 374 GOLONEVTLRNEVAQLKQLLAHKDCPVTAMQKK 408
   | | : | | | | | : | | | : | | : | | : | | : | | : | | : | | :

RESULT 12
Q91576 PRELIMINARY; PRT; 486 AA.
AC Q91576;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE ACTIVATING TRANSCRIPTION FACTOR 2.
GN ATF2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95180723.
RA VILLARREAL X.C., RICHTER J.D.;
RT "Analysis of ATF2 gene expression during early Xenopus laevis
   development."
RL Gene 153:225-229(1995).
DR EMBL; U16158; AAA69518.1; -.
DR HSP; P08047; 1SP2.
DR PFAM; PF00170; bZIP; 1.
DR PFAM; PF00096; zf-C2H2; 1.
SQ SEQUENCE 486 AA; 52378 MW; 4F7C23E3 CRC32;

Query Match          7.0%; Score 135; DB 13; Length 486;
Best Local Similarity 21.5%; Pred. No. 0.1;
Matches 88; Conservative 56; Mismatches 148; Indels 118; Gaps 17;

QY 52 TLKLEPFEED-VLGAEMWESSDLGSFL---DALGDNHERLHPFESNLEFTSLITPDST 107
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 34 TLKFGPARNDSVIVAD--QTPTPTFRFKNCEEVGLFNLASPFEN---EFKASEDDIKK 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 108 VSKDILSTLQFPQPNIPLYASHGAEDFSAETEFENHLSP-PDPSPROV-----APVI 160
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 88 MPLDL-----SPLATPIIRKIEEPSVWETHQD--SPLPHESTTSDQHNQVPLA 136
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```

```

QY 161 N-----LEPVELTASHMTVISPDLGLGMELASESLFTTDLDFVNFDSAVGSIIGGA 212
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 137 QTAOPTSAIVRPASLPVNPVLLTSSDSSVLIQQAIPSTSTVITQAPSSNRPIVPGFP 196
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 213 EELL-----GSPLSVDD-----VESTISFGSPSPSPETSQS 244
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 197 FVLLHLPSGQTMVPAIPASITNSNVHVPAAVPLLRPLTWMVSPGIPGSPSPQVQS-- 254
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 245 IESSPELYKVISSIDASKRFS-----PYSRSKSKSQSVKT 281
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 255 -----EAKMRFNAATQQHPVTVNGDTVKGHSGFAITQSEESRPSLO 298
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 282 SDAKAPKTR-TPAQVPE-----HVIMEHLDDKDRKKLQ-KNNAAIRYRMKKKG 330
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 299 QPATSTETRPSPAQPTQSTSGRRRAANEDPDEKRSKLIQNRRAAASRCQKRVK 358
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 331 AQGTGGEQLEELNTKLTKYVDDDLQREIKYMKNLK--EDVCCKAKGIQK 378
   | | : | | | | | : | | | : | | : | | : | | : | | : | | : | | :
Db 359 VQSLKKADELISLNGYLQNEVTLRNEVAQLKQLLAHKDCPVTAMQKK 408
   | | : | | | | | : | | | : | | : | | : | | : | | : | | : | | :

RESULT 13
Q62870 PRELIMINARY; PRT; 487 AA.
AC Q62870;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE CAMP RESPONSE ELEMENT BINDING PROTEIN 1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-SPRAGUE DAWLEY; TISSUE-BRAIN;
RA MURAMATSU S.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; U38938; AAA93263.1; -.
DR HSP; P08047; 1SP2.
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; bZIP; 1.
DR PFAM; PF00096; zf-C2H2; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 487 AA; 52286 MW; 29CEDF55 CRC32;

Query Match          7.0%; Score 134; DB 11; Length 487;
Best Local Similarity 21.9%; Pred. No. 0.12;
Matches 86; Conservative 58; Mismatches 166; Indels 82; Gaps 16;

QY 52 TLKLEPFEED-VLGAEMWESSDLGSFL---DALGDNHERLHPFESNLEFTSLITPDST 107
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 34 TLKFGPARNDSVIVAD--QTPTPTFRFKNCEEVGLFNLASPFEN---EFKASEDDIKK 88
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 108 VSKDILSTLQFPQPNIPLYASHGAEDFSAETEFENHLSP-PDS---PEQVAPVIN-- 161
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 89 MPLDL-----SPLATPIIRKIEEPSVWETHQDSPLPHESTNDKEIPLAOTA 139
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 162 -----LEPVELTASHMTVISPDLGLGMELASESLFTTDLDFVNFDSAVGSIIGGAEL 215
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 140 QPTSAIVRPASLPVNPVLLTSSDSSVLIQQAIPSTSTVITQAPSSNRPIVPGFPPL 199
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 216 L-----GSPLSVDD-----VESTISFGSPSPSPETSQS 247
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 200 LLHLPNGQTMVPAIPASITNSNVHVPAAVPLVRPVTWVPSPGIPGSPSPQVQS---EA 256
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 248 SPELYKVIS-----TSSIDASKRFSYRSKSKSQSVKTSDAKAPKTRTPAQ--VP 298
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 257 KMLKAALTOQHPPVTVNGDTVKGHSGVLRQAQSESRPQSLQQAIPATSTETPASPAHTTP 316
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 299 E-----HVIMEHLDDKDRKKLQ-KNNAAIRYRMKKKGAEQIKGEQEELNKL 348
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```

Db 317 QTONTSGRRRAANDPDEKRRKFLERNRAASRCQRKRVVOSLEKKAEDLSLNGOL 376
Qy 349 KTKVDLQREIKYMKNL--EDVCKAKGIQK 378
Db 377 QSEVTLRNEVAQLKQLLAHKDCPVTAMQK 408
RESULT 14
Q64089 PRELIMINARY; PRT; 448 AA.
AC Q64089;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE ACTIVATING TRANSCRIPTION FACTOR 2
DE (CYCLIC AMP RESPONSE ELEMENT DNA-BINDING PROTEIN ISOFORM 3).
GN ATF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92123199.
RA GEORGOPOULOS K., MORGAN B.A., MOORE D.D.;
RT "Functionally distinct isoforms of the CRE-BP DNA-binding protein
mediate activity of a T-cell-specific enhancer.";
RL Mol. Cell. Biol. 12:747-757(1992).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL: S76655; AAB21127.1; -.
DR HSSP: P03069; 2DGC.
DR MGD: MGI:109349; Atf2.
DR PROSITE: PS00036; BZIP_BASIC; 1.
DR PFAM: PF00170; bzip; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE .448 AA; 48267 MW; 6698E12B CRC32;
Query Match 6.9%; Score 133; DB 11; Length 448;
Best Local Similarity 21.0%; Pred. No. 0.12;
Matches 76; Conservative 53; Mismatches 157; Indels 76; Gaps 12;
Qy 78 DALGDNHERLHPFESNLEFTSLTPDDSTVSKDILSSTLQFTQPVNPIYASHGAEDF 137
Db 22 EEVGLFENELASPFEN--EFKKASEDDIKMPLDL-----SPLATPIIRSKIEEPS 69
Qy 138 SAETEFEN-----HLSPPDSPQVAPVIN-----LEPVELTASHMTVISPDGLLGME 185
Db 70 VVEITHQDSPLPHESTTSDEKEVPLAQTAQTSIAIVRPASLQVPNVLLTSSDSVITQQ 129
Qy 186 LASESLTFELDFVNFNDSAVSGIGGABELL-----GSPLSVDD----- 224
Db 130 AVPSPTSSTVITQAPSSNRPIVPVGPFPPLLHLPNGQTMPVAIPASITSSNVHVPAAVP 189
Qy 225 -----VESTISFGSPSPSTSSSIIESPELYKVIS-----TSSIDAKRFPYSVR 271
Db 190 LVRPVTWVPSPVGPSPSPQVQS---EAKWLKAALTQOHPVNTGTVKGHSGGLVR 246
Qy 272 SSKSKQSVKTSDAKAPKTRTPAQP---VPE-----HVIMEHDKDKRKKLO--NKN 318
Db 247 TQSESRPQSLQQQATSTSTETPASPAPHTPQTQNTSGRRRAANEDDPDEKRRKFLERNRA 306
Qy 319 AAIYRMRKKGEAGCIKEGEELEELNFKLTKYDDQLQREIKYMKNL--EDVCKAKGIQ 376
Db 307 AASRCRQKRVVOSLEKKAEDLSLNGQLQSEVTLRNEVAQLKQLLAHKDCPVTAMQ 366
Qy 377 LK 378
Db 367 KK 368
RESULT 15

O08592 PRELIMINARY; PRT; 2390 AA.
AC O08592;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE VERSICAN V0 ISOFORM (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR KYOTO;
RA LEMIRE J.N., BRAUN K.R., MAUREL P., MARGOLIS R.U., SCHWARTZ S.M.,
RA WIGHT T.N.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR KYOTO;
RA MILEV P., MAUREL P., CHIBA A., MEVISSSEN M., POPP S., YAMAGUCHI Y.,
RA MARGOLIS R.K., MARGOLIS R.U.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 2073-2115 FROM N.A.
RC TISSUE=NORMAL KIDNEY;
RA PYKE C., KRISTENSEN P., OSTERGAARD P.B., OTURAI P.S., ROMER J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF062402; AAC40166.1; -.
DR EMBL: U75306; AAB51125.1; -.
DR HSSP: P01132; IEPH.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PFAM: PF00008; EGF; 2.
DR PFAM: PF00059; Lectin_c; 1.
DR PFAM: PF00084; sushi; 1.
KW Glycoprotein; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 2390 AA; 261150 MW; DBE8E4A4 CRC32;
Query Match 6.8%; Score 131; DB 11; Length 2390;
Best Local Similarity 21.4%; Pred. No. 1.5; Indels 102; Gaps 15;
Matches 82; Conservative 53; Mismatches 147;
Qy 20 MPVVQTDQFGDLKTSRRHGDESLQPO--GATLKLEPFEDVLGAENMESSDLGSELD 78
Db 1597 IPGIQTDLD--SDIPLGSHSSEESLEVEQKEYATINLSPTEAFDGS-----D 1644
Qy 79 ALGDNHERLHPFESNLEFTSLITPDSTVSKDILSSTLQFTQPVNPIYASHGAEDFS 138
Db 1645 ALPAGHTQAIYNES-----VTPSDGKQPEDI---SFSFAT---GIPV-----SS 1682
Qy 139 AETEFENHLSPPDSPQVAPVINLEPVELTASHMTVISPDGLLGMEELASESLTETLDF 198
Db 1683 TETEL--NTFFPVTSLHIPSKLTTASPEIDKPNIBALSDDIF-----ESTSLSDGA 1734
Qy 199 VNFNDSAVSGIGGABELLGSPLSDVDVESTISFGSPS----- 235
Db 1735 IADQSEVISTGLHLEK-----TQEEYEK--KYGFSQPEFFSGVGEVFTDAPYVSG 1787
Qy 236 -----SPETSOSSIIESPELYKVIS--SSIDAKRFPYSRKSQSVKTSDAKAPRK 289
Db 1788 RTYSVAQPLTEFPNVVQSDSTHYTEATSAVSVTELSPTSPSPSPVYIDSGVSEFTEV 1847
Qy 290 TETPAQVPEHVIMEHDKDKRKKLQNKNAIYRMRKKGE-----AQGIK 335
Db 1848 PHKSAQAPTAASSQKLIEGSEFKYR---ANIEATIKSLGENDHGTSPSMSPALDIS 1904
Qy 336 GEE-----QEELELNTLTKTKVD 353
Db 1905 EDDSKPKLLEDLETSTPTTETSQD 1928

Search completed: August 8, 2000, 03:26:40
Job time: 6191 sec

gb_est9:AA595476	-	200.50	349.85	1.2e-10	582	! AA595476	no34f11.s1	NCI_CGAP
gb_est38:AW513455 <td>-</td> <td>200.50</td> <td>349.45</td> <td>1.2e-10</td> <td>603</td> <td>! AW513455</td> <td>x08ac01.x1</td> <td>NCI_CGAP</td>	-	200.50	349.45	1.2e-10	603	! AW513455	x08ac01.x1	NCI_CGAP
gb_est38:AW170467 <td>-</td> <td>200.50</td> <td>348.57</td> <td>1.4e-10</td> <td>651</td> <td>! AW170467</td> <td>xn62a12.x1</td> <td>Soares_N</td>	-	200.50	348.57	1.4e-10	651	! AW170467	xn62a12.x1	Soares_N
gb_est36:AW001098 <td>-</td> <td>200.50</td> <td>348.25</td> <td>1.4e-10</td> <td>670</td> <td>! AW001098</td> <td>wu24c01.x1</td> <td>Soares_D</td>	-	200.50	348.25	1.4e-10	670	! AW001098	wu24c01.x1	Soares_D

seq_documentation_block:
LOCUS AI479769 769 bp mRNA EST 14-APR-1999
DEFINITION tm69809.x1 NCI_CGAP_Brn25 Homo sapiens CDNA clone IMAGE:2163553
similar to gb:D90209 DNA-BINDING PROTEIN TAXREB67 (HUMAN);, mRNA
sequence.
ACCESSION AI479769
VERSION AI479769.1 GI:4372937
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo

REFERENCE
1 (PAGES 1 TO 69)
NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS
National Cancer Institute / National Institute of Neurological
TITLE
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
JOURNAL
Unpublished (1998)
COMMENT
On Jan 19, 1998 this sequence version replaced gi:2286371.
Contact: Robert Strausberg, Ph.D.

FEATURES
source 1...769
Location/Qualifiers
High quality sequence stop: 451.
Seq primer: -400P from Gibco
Insert length: 334 Std Error: 0.00
www-bio.llnl.gov/bbrp/image/image.html
found through the I.M.A.G.E. Consortium/LLNL at:
Clone distribution: NCI-CGAP clone distribution Information can be
DNA sequencing by: Washington University Genome Sequencing Center
CDNA Library Arrayed by: Greg Lennon, Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
Ph.D.
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.

```

FEATURES
source
high quality sequence top: 431.
Location/Qualifiers
1. .769
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2163353"
/clone_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pTT3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTCACATCTCAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pTT3 vector.
Library is normalized, and was constructed by Banto

```

BASE COUNT	162 a	203 c	189 g	214 t	1 others
ORIGIN	Soares and M.Fatima Bonaldo."				

```
alignment_scores:
  Quality: 218.00      Length: 239
  Ratio: 1.690        Gaps: 7
  Percent Similarity: 53.975  Percent Identity: 28.870
```

alignment block:

US-08-656-811A-1 x A1479769/rev ..
Align seg 1/1 to reverse of: A1479769 from: 1 to: 769

152 SerProGluGlnValAlaProValIleAsnLeuGluProValGluLeuTh 168

152 SerProGluGlnValAlaProValIleAsnLeuGluProValGluLeuTh 168


```

231 PheSerGlyProSerSerProGluThrSerGlnSerSerIleIleGluSe 247
498 .....ACCCCTTCAGATAATGATAGTGCATCTGT...AT 467
247 rSerProGluLeuTyrLysValIleSerThrSerSerIleAspAlaSerL 264
466 GAGCCAGAGTCTCTATG.....GGGTCTC 441
264 ysArgPheSerProTyrSerArgSerSerLysSerLysGlnSerValLys 280
440 CTCAGCACAGCCCTCTACCAAGGGCTCTCCAAATAGG...AGCCTCCCA 394
281 ThrSerAspAlaLysAlaProArgLysThrArgThrProAlaGlnProVa 297
393 TCTCCAGTGTCTCTGTGGGTCTGCCGCTCCCAACCTTACGATCTCC 344
297 lProGluHisValIleMet.....GluHisLeuAspLysL 309
343 TGGAGAGAAGATGGTAGCAGCAAAAGTAAGGCTCAGAACTGGTAAGA 294
309 ysAspArgLysLysLeuGlnAsnLysAsnAlaIleArgTyrArgMet 325
293 AGCTGAAAAAATGGACAAAACAGACAGCAGCCACTAGGTACGCCAG 244
326 LysLysLysGlyGluAlaGlnGlnLysLysGlyGluGlnGlnGluLeuG 342
243 AAGAAGGGGGGAGCAGAGGCTCTTACTGTGTGAGTGCACAAAGAGCTG 194
342 uGluLeuAsnThrLysLysLeuLysThrLysValAspLysLeuGlnArg 359
193 AAGAAGNACGAGGCTCTAAAGAGAGAGGGCGGATCTCTGCCAAGGAGA 144
359 leLysTyrMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
143 TCCAGTACCTCAAGATTGTAGAGAGGTCGCCAAGGCAAGGGGG 97

seq_name: gb_est16:A1114565

seq_documentation_block:
LOCUS A1114565 652 bp mRNA EST 11-NOV-1999
DEFINITION H1149 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
sequence.
ACCESSION A1114565
VERSION A1114565.1 GI:6359910
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Yu, Y., Zhang, C., Luo, L., Ouyang, S., Li, W., Wu, J., Zhou, S., Liu, M.
and He, F.
TITLE Expression profile analysis of a human fetal liver cDNA library
JOURNAL Unpublished (1998),
COMMENT Contact: Rongtao Yu
Department of Hematology
Beijing Institute of Radiation Medicine
27 Taiping Road, Beijing 100850, P.R.China
Tel: 0086-10-68159479
Fax: 0086-10-68214653
Email: yyt48@yahoo.com.
FEATURES
Location/Qualifiers
1..652
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human fetal liver cDNA library"
/dev_stage="fetal"
/lab_host="MC1061/P3"
/note="Vector: pCDNA1"
BASE COUNT 204 a 137 c 177 g 128 t 6 others
ORIGIN

```

```

alignment_scores:
Quality: 213.00 Length: 224
Ratio: 1.805 Gaps: 7
Percent Similarity: 52.679 Percent Identity: 31.696

alignment_block:
US-08-656-811A-1 x A1114565 ..
Align seg 1/1 to: A1114565 from: 1 to: 652

158 ProValIleAsnLeuGluPro.....ValGluLeu 167
|||||:..... |||
13 CCAGTACTGCAGACTATGTCATGTTCAAGTGGGACCTCAAGCTGCACAAG 62
167 urThAlaSerHisMetThrValIleSerProAspGlyLeuLeuGlyGly 184
||||| ||||| ||| :|||:|||||
63 TACCCGCTTAACATATG.....CCTACTTCGCTGGTGGGATGC 100
184 etGluLeuAlaSerGluSerLeuThrPheThrGluLeuAspPheValAsn 200
:||||: ||| ||| :|||
101 TGGAGATGCCTTTGATCCCTCAGTCATTAAGGAGGAGACACACCCCTTCA 150
201 PheAsnAspSerAlaValGlySerIleGlyGlyAlaGluGluLeuLeuG 217
||||| ||| :|||: |||||
151 GATAATGATAGT.....GGCATCTGTATGAGCCAGAGTCTCTATCTGGG 194
217 ySerProLeuSerValAspValGluSerThrIleSerPheSerGlyP 234
|||||
195 GTCCTCT.....C 202
234 rSerSerProGluThrSerGlnSerSerIleIleGluSerSerProGlu 250
|||||: ||| :||| |||
203 AGCACAGCCCTTACC...AGGGCTCTCCAATAGGACCTCCCA... 246
251 LeuTyrLysValIleSerThrSerSerIleAspAlaSerLysArgPheSe 267
||| :||| :||| ||| :|
247 .....TCTCCAGGTGTCTCTGTGGTCTGCCCTGCCCAA 281
267 rProTyrSerArgSerSerLysSerLysGlnSerValLysThrSerAspA 284
:|||||: ||| :|||: |||: |||: |||: |||: |||: |||: |||
282 ACCTTACGATCCTCTCGAGAGAAAGATGTCAGCAAAAGATAAAGGTT. 330
284 laLysAlaProArgLysThrArgThrProAlaGlnProValProGluHis 300
330 ..... 330
301 ValIleMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsnLy 317
||| ||||| ||||| :||| |||||
331 .....GAGAAACTGCATAAAGAGCTGAAAAAATAATGGACAAACAA 371
317 sAsnAlaAlaIleArgTyrArgMetLysLysLysGlyGlyAlaGlnGlyI 334
|:||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
372 GACAGACCCACTAGTACCCGCCAAGAAGAGGGCGGAGAGAGGCTC 421
334 leLysGlyGluGlnGlnGluLeuGluLeuGluLeuAsnThrLysLysThr 350
: ||||| :|||: |||||: ||| |||||
422 TTACTGGTGGTGCAGAGAGCTGGAAAGAACAGACGAGGCTCTAAAGAG 471
351 LysValAspAspLeuGlnArgGluIleLysTyrMetLysAsnLeuMetG 367
:|||||: ||| :|||: |||: |||: |||: |||: |||: |||: |||
472 AGGGCGGATTCCTCGCCCAAGGAGATCCAGTACCTGAAAGATTGATAGA 521
367 uAspValCysLysAlaLysGly 374
|:||||| |||||:|||||
522 AGAAGTCCGCAAGGCAAGGGGG 543

seq_name: gb_est22:A1598953

seq_documentation_block:
LOCUS A1598953 577 bp mRNA EST 21-APR-1999
DEFINITION REMD233 3' end, mRNA sequence.

```

```

ACCESSION   AI598953
VERSION     AI598953.1  GI:4608001
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 577)
AUTHORS     Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE       Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
JOURNAL
COMMENT     Unpublished (1998)
            On Jan 19, 1998 this sequence version replaced gi:2284562.
            Other_ESTs: TC74232
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES
            source
            1..577
            /organism="Rattus sp."
            /db_xref="taxon:10118"
            /clone="REM233"
            /clone_lib="Normalized rat embryo, Bento Soares"
            /dev_stage="embryo 8, 12, 18 dpc"
            /note="Vector: pT73Pac; Site_1: ECORI; Site_2: NotI"
BASE COUNT   132 a 151 c 124 g 170 t
ORIGIN
alignment_scores:
Quality: 211.00 Length: 156
Ratio: 2.153 Gaps: 4
Percent Similarity: 62.821 Percent Identity: 37.821
alignment_block:
US-08-656-811A-1 x AI598953/rev ..
Align seg 1/1 to reverse of: AI598953 from: 1 to: 577
233 GlyProSerSerProGluThrSerGlnSerSerIlelleGluSerSerPr 249
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560 GGAGATAGGAAGCGCTGACTCT.....GCTGCTTATATTACTCTAACCC 517
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
249 oGluLeuTyrlsValIleSerThr...SerSerIleAspAlaSerLysA 265
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
516 TCAGTGTGAAGGAGGAGGACATCCCTCTGATAGTGACAGTGGCATCT 467
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 rgPheSerProTySerArgSerLysSerLysGlnSerVallylsth 281
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
466 GTATGAGCCCTGAGTCTCTACCTACCTGGCTCTCCCAACACAGCCCTCCACC 417
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 SerAspAlaLys.....AlaProArgLysTh 290
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
416 TCCAGGGCCCCACCAGACAGTCTCCCTTCTCCAGGTGTCTCGTGGTTC 367
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
290 rArg.....ThrProAlaGlnProValProGluHisValI 302
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
366 TCGACCCCAACCTTATGACCCACCTGGAGTAGTGTGACAGCTAAAGTGA 317
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 leMetGluHisLeuAspLysLysAspArgLysLysLeuGlnAsnLysAsn 318
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
316 AGACTGAAAGTGTGATAAGAAGCTGAAAGAAGATGGACAAACAAGACA 267
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
319 AlaAlaIleArgTyArgMetLysLysLysGlyGluAlaGlnGlyIleLy 335
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
266 GCACCTACTAGTACCGCCAGAGAAGAGGCGTGCAGGAGGCCCTCCAC 217

```

```

335 sGlyGluGlnGlnGluLeuGluLeuAsnThrLysLeuLysThrLysV 352
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 TGGCGAGTGTAAAGAGCTAGAAAGAAGACGAGCGCTCTGAAGAAGAGG 167
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352 alAspAspLeuGlnArgGluLeuLysTyMetLysAsnLeuMetGluAsp 368
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166 CAGATTCTCTCCCAAGAGATTTCAGTATCTAAAGACCTGATAGAAGAG 117
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
369 ValCysLysAlaLysGly 374
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 GTCCGTAAAGCAAGGGG 99
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: gb_est16:AI101524
seq_documentation_block:
LOCUS     AI101524          582 bp      mRNA          31-JAN-1999
DEFINITION EST210813 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
            RBRBR68 3' end, mRNA sequence.
ACCESSION AI101524
VERSION   AI101524
KEYWORDS  EST.
SOURCE    Rattus sp.
ORGANISM Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE 1 (bases 1 to 582)
AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
JOURNAL   Unpublished (1998)
COMMENT   Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES
            source
            1..582
            /organism="Rattus sp."
            /db_xref="ATCC (inhost):2024131"
            /db_xref="taxon:10118"
            /clone="RBRBR68"
            /clone_lib="Normalized rat brain, Bento Soares"
            /note="Organ: brain; Vector: pT73Pac; Site_1: ECORI;
            Site_2: NotI"
BASE COUNT   133 a 153 c 124 g 172 t
ORIGIN
alignment_scores:
Quality: 211.00 Length: 156
Ratio: 2.153 Gaps: 4
Percent Similarity: 62.821 Percent Identity: 37.821
alignment_block:
US-08-656-811A-1 x AI101524/rev ..
Align seg 1/1 to reverse of: AI101524 from: 1 to: 582
233 GlyProSerSerProGluThrSerGlnSerSerIlelleGluSerSerPr 249
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560 GGAGATAGGAAGCGCTGACTCT.....GCTGCTTATATTACTCTAACCC 517
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
249 oGluLeuTyrlsValIleSerThr...SerSerIleAspAlaSerLysA 265
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
516 TCAGTGTGAAGGAGGAGGACATCCCTCTGATAGTGACAGTGGCATCT 467
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 rgPheSerProTySerArgSerLysSerLysGlnSerVallylsth 281

```

```

150 ProAspSerPro.....GluInValAlaProValIleAsnLeuGI 163
||||:||||| ::: :|||||: ::: :|||
23 CCTGAATCTCCAATTGGAGCAGATCCCATGGCCCTTTATCTTCGCTCTG 72
163 uProValGluLeuThrAlaSerHisMetThrValIleSerProAspGlyL 180
||||: ::: :|||| ::: :|||||
73 GTCCTTTCCCTCTCCCCAGGGCTCTGACTCTCTGTT...CCAGCACCATT 119
180 euLeuGlyGlyMetCluLeuAlaSerGlu..... 189
: : : : :
120 CATTT...AGTTTAGAAGTAGGTAGTGAAGTGGATGTTCTGGAAGGTGAA 166
190 .....SerLeuThrPheThrGluLe 196
167 AGGAACAGGAGGAGGCCCCACCTTTCTGCTAGTGTGATCACCAGCTCAGAGAA 216
196 uAspPheValAsnPheAsnAspSerAlaValGlySerIleGlyGlyAlaG 213
: : : : : : : : : : : : : : : :
217 AGAGGAGGAGGAACCATTCGCGATGATAGT...GGAATATGCATGAGCCCCAG 263
213 luGluLeuLeuGlySerProLeuSerValAspValGluSerThrIle 229
: : : : : |||||: :|||
264 ACCTTATTTGGGAACACC..... 283
230 SerPheSerGlyProSerSerProGluThrSerGlnSerSerIleIleGI 246
||||| : : : : : : : : : : : : : : :
284 CAACATAGCTCCATACCAATTCATTCATTCGATCACCACCAATCA 321

```

105062WY T HTTQ 000 80 EN

KEYWORDS

VERSION AA903759.1 GI:3038882

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 849)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Feb 6, 1998 this sequence version replaced gi:2843818.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbrrp/image/image.html
 Insert Length: 1059 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 308.
 Location/Qualifiers
 1. .849
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1518894"
 /clone_lib="NCI-CGAP_GC4"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH108"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; 1st strand cDNA was prepared from 3 pooled
 germ cell tumors, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT73
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bonaldo."
 197 a 216 c 205 g 231 t
 BASE COUNT
 ORIGIN
 alignment_scores:
 Quality: 209.00 Length: 250
 Ratio: 1.504 Gaps: 11
 Percent Similarity: 55.600 Percent Identity: 30.000
 alignment_block:
 US-08-656-811A-1 x AA903759/rev ...
 Align seg 1/1 to reverse of: AA903759 from: 1 to: 849
 143 PheGluAsnHisLeuSerProAspSerProGluGlnValAlaProVa 159
 757 TTCGGGGAACTATTAGGGCCCC.....CCCAAAAGGGTAACCCCAAT 714
 159 lileAsnLeuGluProValGluLeuThrAlaSerHisMetThrVal 176
 713 T.....GCCCTTT 706
 176 erProAspGlyLeuLeuGlyGlyMetGluLeuAlaSerGluSerLeuThr 192
 705 CCCGAAGGTTTACAAACCCGACAGGTTGCC.....CCTTTAACT 662
 193 PheThrGluLeuAspPheValAsnPheAsnAspSerAlaValGlySerIl 209
 661 TTCCTAAAA...ACTTTTCCCTTTTCCAGGGGCTCTTGCTTCCACTC 615
 209 eGlyGlyAlaGluLeuLeu...LeuGlySerProLeuSerValAspAsp 224
 614 CGGATCATTCCTTTAGTTTACAGTTGGGCGAGTGAAGTGGATATCACTCA 565

225 ValGluSerThrIleSerPheSerGly..... 233
 564 GGATAGGAAGCAGACTACACTGCTTACCTGCTCCATGATCCCTCAGTG 515
 234ProSerSerProGluThrSerGlnSerSerIleIleGluS 247
 514 CATAAAGGAGGAGACACCCCTTCAGATAATGATAGTTGCATCTGT...A 468
 247 erSerProGluLeuThrLysValIleSerThrSerSerIleAspAlaSer 263
 467 TGAGCCCAAGAGTCTCTATCTG.....GGGTCT 442
 264 LysArgPheSerProTyrSerArgSerSerLysSerLysGlnSerValLy 280
 441 CCTCAGCACAGCCCTCTACAGGGGCTCTCCAAATAGG...AGCCTCCC 395
 280 sThrSerAspAlaLysAlaProArgLysThrArgThrProAlaGlnProV 297
 394 ATCTCCAGGTCTTCTCTGTGGTCTGCCCGTCCCAAACTTACGATCCTC 345
 297 alProGluHisValIleMet.....GluHisLeuAspLys 308
 344 CTGAGAGAGAGATGCTAGCAGCAAAAGTAAAGGGTGAGAACTGGATAAG 295
 309 LysAspArgLysLysLeuGlnAsnLysAsnAlaAlaIleArgTyrArgMe 325
 294 AAGTGAAAAAATGGAGCAAAACACAGACAGCAGCAGCTAGTACCGCCA 245
 325 TlysLysLysGlyGluAlaGlnGlyIleLysGlyGluGlnGluLeuG 342
 244 GAAGAAGGGCGGAGCAGGAGGCTCTTACTGCTGAGTGCAAGAGCTGG 195
 342 luGluLeuAsnThrLysLeuLysValAspAspLeuGlnArgGlu 358
 194 AAAAGAGAAGCAGGCTCTAAAGAGAGAGGGCGGATTCCTGGCCCAAGG 145
 359 IleLysTyrMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
 144 ATCCAGTACCTGAAGATTGATAGAGAGGTCGCCAAGGCAAGGGGG 97

seq_name: gb_est18:AI237099

seq_documentation_block:
 LOCUS AI237099 545 bp mRNA EST 31-JAN-1999
 DEFINITION EST233661 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
 ROVD064 3' end, mRNA sequence.
 ACCESSION AI237099
 VERSION AI237099.1 GI:3830605
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 545)
 AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kerlavage,A.R. and Adams,M.D.
 TITLE Rat Genome Project: generation of a Rat EST (REST) Catalog & Rat
 Gene Index
 JOURNAL Unpublished (1998)
 COMMENT On Apr 7, 1998 this sequence version replaced gi:3036733.
 Other ESTs: TC61418
 Contact: Lee, NH
 ATCC
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: MJ3-21.
 Location/Qualifiers
 1. .545
 FEATURES
 source

```

/organism="Rattus sp."
/db_xref="ATCC (inhost):2042258"
/db_xref="taxon:10118"
/clone="ROVDO64"
/clone_lib="Normalized rat ovary, Bento Soares"
/notes="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"

```

```

BASE COUNT      124 a 141 c 120 g 160 t
ORIGIN

```

```

alignment_scores:
  Quality: 208.00      Length: 147
  Ratio: 2.237         Gaps: 3
  Percent Similarity: 63.265      Percent Identity: 38.776

```

```

alignment_block:
US-08-656-811A-1 x AI237099/rev ..

```

```

Align seg 1/1 to reverse of: AI237099 from: 1 to: 545

```

```

242 SerSerIleIleGluSerProGluLeuTyrlYsValIleSerThr... 257
      : : : : : ||| : : : : : ||| : : : : :
539 GCTGCTTATATTACTCTAACCCCTCAGTGTCTAAAGGAGGACACTCC 490
      : : : : : ||| : : : : : ||| : : : : :
258 .SerSerIleAspAlaSerLysArgPheSerProTyrlSerArgSerL 274
      : : : : : ||| : : : : : ||| : : : : :
489 CTCGTAGTGTGACAGTGCATCTGTATGAGCCCTGAGTCTACCTGGGCT 440
      : : : : : ||| : : : : : ||| : : : : :
274 ySSerLysGlnSerVallyThrSerAspAlaLys..... 285
      : : : : : ||| : : : : : ||| : : : : :
439 CTCGCCAACACAGCCCTTCCACCTCCAGGCCCCACCACGAGTCTGGCT 390
      : : : : : ||| : : : : : ||| : : : : :
286 .....AlaProArgLysThrArg.....ThrProAl 294
      : : : : : ||| : : : : : ||| : : : : :
389 TCTCCAGGTGTCTCTCGTGGTTCGACCCCAACCTTATGACCCACCTGG 340
      : : : : : ||| : : : : : ||| : : : : :
294 aGlnProValProGluHisValIleMetGluHisLeuAspLysAspA 311
      : : : : : ||| : : : : : ||| : : : : :
339 AGTTAGTGTGACAGCTAAAGTGAAGACTGAAAGTTGGATAAGAGCTGA 290
      : : : : : ||| : : : : : ||| : : : : :
311 rGlyLysLeuGlnAsnLysAsnAlaAlaIleArgTyrlArgMetLys 327
      : : : : : ||| : : : : : ||| : : : : :
289 AAAAGATGGAGCAAAACAAAGACACAGCTACTAGGTACCCGCAAGAA 240
      : : : : : ||| : : : : : ||| : : : : :
328 LysGlyGluAlaGlnGlyIleLysGlyGluGlnGluLeuGluLe 344
      : : : : : ||| : : : : : ||| : : : : :
239 AGGCTGAGCAGGAGGCCCTCAGTGGGAGTGTAAAGAGCTAGAAAGAA 190
      : : : : : ||| : : : : : ||| : : : : :
344 uAsnThrLysLeuLysThrLysValAspAspLeuGlnArgGluLeu 361
      : : : : : ||| : : : : : ||| : : : : :
189 GAACGAGGCTCTGAAGAGAGAGGAGGAGTCTCTGCCCAAGAGATTCA 140
      : : : : : ||| : : : : : ||| : : : : :
361 yrMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
      : : : : : ||| : : : : : ||| : : : : :
139 ATCTAAAGACCTGATAGAAGAGGTCCGTAAGGCAAGGGGG 99
      : : : : : ||| : : : : : ||| : : : : :

```

```

seq_name: gb_est20:AI408946

```

```

seq_documentation_block:
LOCUS      AI408946      549 bp      mRNA      EST      09-FEB-1999
DEFINITION EST237237 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
            RKIDR39 3' end, mRNA sequence.
ACCESSION  AI408946
VERSION    AI408946
KEYWORDS   EST.
SOURCE     Rattus sp.
ORGANISM   Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 549)
AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,

```

```

seq_name: gb_est14:AA946061

```

```

TITLE      Kerlavage,A.R. and Adams,M.D.
            Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
JOURNAL    Unpublished (1998)
COMMENT    On Jan 19, 1998 this sequence version replaced gi:2286321.
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES   Location/Qualifiers
            source
              1..549
              /organism="Rattus sp."
              /db_xref="taxon:10118"
              /clone="RKIDR39"
              /clone_lib="Normalized rat kidney, Bento Soares"
              /notes="Organ: kidney; Vector: pT7T3Pac; Site_1: EcoRI;
              Site_2: NotI"
BASE COUNT      126 a 142 c 121 g 160 t
ORIGIN

```

```

alignment_scores:
  Quality: 208.00      Length: 147
  Ratio: 2.237         Gaps: 3
  Percent Similarity: 63.265      Percent Identity: 38.776

```

```

alignment_block:
US-08-656-811A-1 x AI408946/rev ..

```

```

Align seg 1/1 to reverse of: AI408946 from: 1 to: 549

```

```

242 SerSerIleIleGluSerProGluLeuTyrlYsValIleSerThr... 257
      : : : : : ||| : : : : : ||| : : : : :
539 GCTGCTTATATTACTCTAACCCCTCAGTGTCTAAAGGAGGACACTCC 490
      : : : : : ||| : : : : : ||| : : : : :
258 .SerSerIleAspAlaSerLysArgPheSerProTyrlSerArgSerL 274
      : : : : : ||| : : : : : ||| : : : : :
489 CTCGTAGTGTGACAGTGGCCTGTGTAGCCCTGAGTCTACCTGGGCT 440
      : : : : : ||| : : : : : ||| : : : : :
274 ySSerLysGlnSerVallyThrSerAspAlaLys..... 285
      : : : : : ||| : : : : : ||| : : : : :
439 CTCGCCAACACAGCCCTTCCACCTCCAGGCCCCACCACGAGTCTGGCT 390
      : : : : : ||| : : : : : ||| : : : : :
286 .....AlaProArgLysThrArg.....ThrProAl 294
      : : : : : ||| : : : : : ||| : : : : :
389 TCTCCAGGTGTCTCTCGTGGTTCGACCCCAACCTTATGACCCACCTGG 340
      : : : : : ||| : : : : : ||| : : : : :
294 aGlnProValProGluHisValIleMetGluHisLeuAspLysAspA 311
      : : : : : ||| : : : : : ||| : : : : :
339 ACTTAGTGTGACAGCTAAAGTGAAGACTGAAAGTTGGATAAGAGCTGA 290
      : : : : : ||| : : : : : ||| : : : : :
311 rGlyLysLeuGlnAsnLysAsnAlaAlaIleArgTyrlArgMetLys 327
      : : : : : ||| : : : : : ||| : : : : :
289 AAAAGATGGAGCAAAACAAAGACAGAGCTACTAGTACCCGCAAGAG 240
      : : : : : ||| : : : : : ||| : : : : :
328 LysGlyGluAlaGlnGlyIleLysGlyGluGlnGluLeuGluLe 344
      : : : : : ||| : : : : : ||| : : : : :
239 AGGCTGAGCAGGAGGCCCTCAGTGGGAGTGTAAAGAGCTAGAAAGAA 190
      : : : : : ||| : : : : : ||| : : : : :
344 uAsnThrLysLeuLysThrLysValAspAspLeuGlnArgGluLeu 361
      : : : : : ||| : : : : : ||| : : : : :
189 GAACGAGGCTCTGAAGAGAGAGGAGGAGTCTCTGCCCAAGAGATTCA 140
      : : : : : ||| : : : : : ||| : : : : :
361 yrMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
      : : : : : ||| : : : : : ||| : : : : :
139 ATCTAAAGACCTGATAGAAGAGGTCCGTAAGGCAAGGGGG 99
      : : : : : ||| : : : : : ||| : : : : :

```

```

seq_documentation_block: 551 bp mRNA EST 08-JAN-1999
LOCUS AA946061 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
DEFINITION RLUBA73 3' end, mRNA sequence.
ACCESSION AA946061
VERSION AA946061.1 GI:4132681
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE 1 (bases 1 to 551)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT On May 1, 1998 this sequence version replaced gi:3105977.
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
source 1..551
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone="RLUBA73"
/clone_lib="Normalized rat lung, Bento Soares"
/note="Organ: lung; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 125 a 142 c 121 g 163 t
ORIGIN

alignment_scores:
Quality: 208.00 Length: 147
Ratio: 2.237 Gaps: 3
Percent Similarity: 63.265 Percent Identity: 38.776

alignment_block:
US-08-656-811a-1 x AA946061/rev ..
Align seg 1/1 to reverse of: AA946061 from: 1 to: 551
242 SerSerIleIleGluSerSerProGluLeuTyLysValIleSerThr.. 257
: : : : : ||| : : : : : ||| : : : : :
539 GCTGCTTATATTACTCTACCCCTCAGTGTGTAAGGAGGAGACACTCC 490
: : : : : ||| : : : : : ||| : : : : :
258 .SerSerIleAspAlaSerLysArgPheSerProTySerArgSerL 274
: : : : : ||| : : : : : ||| : : : : :
489 CTCGTAGATGACAGTGCATCTGTATGAGCCCTGAGTCCCTACCTGGGCT 440
: : : : : ||| : : : : : ||| : : : : :
274 ySerLysGlnSerValIleThrSerAspAlaLys..... 285
: : : : : ||| : : : : : ||| : : : : :
439 CTCGCCAACACAGCCCTTCCACCTCCAGGGCCGCCACGAGAGTGCCT 390
286 .....AlaProArgLysThrArg.....ThrProAl 294
: : : : : ||| : : : : : ||| : : : : :
389 TCtCCAGGTGTCCTCTGTTCTCGACCCCAACCTTATGACCCACCTGG 340
294 aGlnProValProGluHisValIleMetGluHisLeuAspLysLysAsp 311
: : : : : ||| : : : : : ||| : : : : :
339 AGTTAGTGTGCACAGCTAAAGTGAAGACTGAAAGCTTGGATAAAGAGCTGA 290
: : : : : ||| : : : : : ||| : : : : :
311 rGlyLysLysLeuGlnAsnLysAsnAlaIleArgTyrArgMetLysLys 327
: : : : : ||| : : : : : ||| : : : : :
289 AAAAGATGGAGCAAAACAGACAGCAGCTACTAGTACCGCCGCAAGAAG 240

```

```

328 LysGlyGluAlaGlnGlyIleLysGlyGluGluGlnGluGluGluLe 344
: : : : : ||| : : : : : ||| : : : : :
239 AGGCTGAGCAGGAAGGCCCTCACTGGCAGTGTAAAGAGCTAGAAAGAA 190
: : : : : ||| : : : : : ||| : : : : :
344 uAsnThrLysLysThrLysValAspAspLeuGlnArgGluIleLysT 361
: : : : : ||| : : : : : ||| : : : : :
189 GAACGAGGCTCTGAAGAGAGAGGAGGATCTCTGCCCAAGAGATTCA 140
: : : : : ||| : : : : : ||| : : : : :
361 yrMetLysAsnLeuMetGluAspValCysLysAlaLysGly 374
: : : : : ||| : : : : : ||| : : : : :
139 ATCTAAGAGACTGATAGAGAGGTCCTTAAGGCAAGGGGG 99
: : : : : ||| : : : : : ||| : : : : :
seq_name: gb_est18:AI237167
seq_documentation_block: 533 bp mRNA EST 31-JAN-1999
LOCUS AI237167 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
DEFINITION ROVDP45 3' end, mRNA sequence.
ACCESSION AI237167
VERSION AI237167.1 GI:3830673
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE 1 (bases 1 to 533)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2285377.
Other_ESTs: TC61418
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
source 1..533
/organism="Rattus sp."
/db_xref="ATCC (inhost):2042319"
/clone="ROVDP45"
/note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 121 a 138 c 118 g 156 t
ORIGIN

alignment_scores:
Quality: 207.00 Length: 144
Ratio: 2.275 Gaps: 3
Percent Similarity: 63.194 Percent Identity: 39.583

alignment_block:
US-08-656-811a-1 x AI237167/rev ..
Align seg 1/1 to reverse of: AI237167 from: 1 to: 533
245 IleGluSerSerProGluLeuTyLysValIleSerThr...SerSerI 260
: : : : : ||| : : : : : ||| : : : : :
525 ATTACTCTAACCCCTCAGTGTGTAAGGAGGAGACACACTCCCTCTGATAG 476
: : : : : ||| : : : : : ||| : : : : :
260 eAspAlaSerLysArgPheSerProTySerArgSerSerLysLysG 277
: : : : : ||| : : : : : ||| : : : : :
475 TGACAGTGGCATCTGTATGAGCCCTGAGTCTACCTGGGCTCTCCCAAC 426

```

```
277 lnSerValLysThrSerAspAlaLys..... 285
285 ACAGCCTTCCACCTCCAGGCCCCACACAGACAGTCTGCCCTTCCACGGT 376
286 AlaProArgLysThrArg.....ThrProAlaGlnProVa 297
375 GTTCTCGTGTTCGACCCCAACCTATGACCCACCTGGAGTAGTGT 326
297 lProGluHisValleMetGluHisLeuAspLysLysAspArgLysLysL 314
325 GACAGCTAAAGTGAAGACTGAAAGTTGGATAAGAGCTGAAAGATGG 276
314 euGlnAsnLysAsnAlaAlaIleArgTyrArgMetLysLysGlyGlu 330
275 AGCAAAACAGACAGACAGCTACTAGGTACCGCCAGAGAGAGCGGTGAG 226
331 AlaGlnGlyLysGlyGluGluGlnGluLeuGluLeuAsnThrLy 347
225 CAGGAGCCCTCAGTGGCGAGTGTAAAGAGCTAGAAAAGAGACGAGGC 176
347 sLeuLysThrLysValAspAspLeuGlnArgGluIleLysTyrMetLysA 364
175 TCTGAAAGAGAGGAGCAGATTCTCTCGCCAAAGAGATTGATCTATAAG 126
364 snLeuMetGluAspValCysLysLysLysGly 374
125 ACCTGATAGAGAGGTCGTAAGCAAGGGGG 94
```

seq_name: gb_est20:AI412959

seq_documentation_block: 540 bp mRNA EST 09-FEB-1999
LOCUS AI412959
DEFINITION EST241259 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
RBRED32 3' end, mRNA sequence.

ACCESSION AI412959

VERSION AI412959.1 GI:4256463

KEYWORDS EST.

SOURCE Rattus sp.

ORGANISM Rattus sp.

REFERENCE 1 (bases 1 to 540)

AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,

Kerlavage,A.R. and Adams,M.D.

TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat

Gene Index

JOURNAL Unpublished (1998)

COMMENT On Sep 25, 1998 this sequence version replaced gi:3658537.

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

Seq primer: M13-21.

Location/Qualifiers

1..540

/organism="Rattus sp."

/db_xref="taxon:10118"

/clone="RBRED32"

/clone_lib="Normalized rat brain, Bento Soares"

/note="Organ: brain; Vector: pT7T3Pac; Site_1: EcoRI;

Site_2: NotI"

BASE COUNT 123 a 140 c 116 g 161 t

ORIGIN

alignment_scores:

Quality: 207.00 Length: 147

Ratio: 2.275 Gaps: 2

Percent Similarity: 61.905 Percent Identity: 37.415

alignment_block:

US-08-656-811A-1 x AI412959/rev ..

Align seg 1/1 to reverse of: AI412959 from: 1 to: 540

```
242 SerSerIleIleGluSerSerProGluLeuTyrLysValIleSerThr.. 257
539 GCTGCTTATTATTACTCTAACCCTCAGTGTGTAAAGAGGAGACACTCC 490
258 SerSerIleAspAlaSerLysArgPheSerProTyrSerArgSerSerL 274
489 CTCTGATAGTACAGTGGCCTCTGTATGAGCCCTGAGTCCTACATGGGCT 440
274 ysSerLysGlnSerValLysThrSerAspAla..... 284
439 CTCCCAACACAGCCCTTCCACCTCCAGGCCCCACACAGACAGTCTGCCT 390
285 .....LysAlaProArgLysThrArgThrProAl 294
389 TCTCCAGGTGTTTCATCGTGGTCTCGACCCCAACCTTATGACCCACCTGG 340
294 agLnpProValProGluHisValIleMetGluHisLeuAspLysLysAspA 311
339 ACTTAGTGTGACAGCTAAAGTGAAGACTGAAAAGTTGGATAAGAAGCTCA 290
311 rgLysLysLeuGlnAsnLysAsnAlaAlaIleArgTyrArgMetLysLys 327
289 AAAAGATGGAGCAAAACACAGACAGCAGCTACTAGTACCGCCAGAGAAG 240
328 LysGlyGluAlaGlnGlyLysGlyGluGlnGluLeuGluGluLeu 344
239 AGGCTGAGCAGGAGGAGCCCTCAGTGGCAGTGTAAAGAGCTAGAAAAGAA 190
344 uasnThrLysLysLysThrLysValAspLeuGlnArgGluIleLysT 361
189 GAACGAGGCTCTGAAAGAGAGAGGAGGAGATTCTCTCGCCAAAGAGATTCA 140
361 yrMetLysAsnLeuMetGluAspValCysLysLysLysLysGly 374
139 ATCTAAAGACCTGATAGAGAGGTCCTGTAAGGCAAGGGGG 99
```

seq_name: gb_est23:AI649113

seq_documentation_block: 737 bp mRNA EST 30-APR-1999

LOCUS AI649113

DEFINITION uk25R03.x1 Sugano mouse embryo mewa Mus musculus cDNA clone

IMAGE:1970069 3' similar to gb:D90209 DNA-BINDING PROTEIN TAXREB67

(HUMAN); gb:M94087 M.musculus mATF4 (MOUSE);, mRNA sequence.

ACCESSION AI649113

VERSION AI649113.1 GI:4729947

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 737)

AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,

Perron,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,

Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,

Waterston,R. and Wilson,R.

TITLE The WashU-NCI Mouse EST Project 1999

JOURNAL Unpublished (1999)

COMMENT On Feb 18, 1999 this sequence version replaced gi:4299336.

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.


```
133 GCATAGCCCTCCACCTCCAGG.....GCCCCACCAGAC. 166
251 euYrLysValIleSerThrSerSerIleAspAlaSerLysArgPheSer 267
   ::::: ||| ::::: ::::: ||| :::
167 .....ATCTGCCTTCACCAGGTGGTCCCGTGGGTCTCCTCGGCCCAA 211
268 ProTyrSerArgSerLysSerLysGlnSerValLysThrSerAspAl 284
   |||||:::
212 CCTATGAC..... 220
284 aLysAlaProArgLysThrArgThrProAlaGlnProValProGluHisV 301
   |||||:::
221 .....CCACCTGGAGTTAGTTTGACAGCTAAAG 248
301 alIleMetGluHisLeuAspLysLysArgLysLysLeuGlnAsnLys 317
   ||| ||||| ||||| ::::: |||||
249 TGAAGACTGGAATTTGNTAAGAGCTGAANAAGATGGAGCAAAACAAG 298
318 AsnAlaIleArgTyrArgMetLysLysLysGlyGluAlaGlnGlyI 334
   ::||| ||||| ::||| ::||| ::||| ::||| ::|||
299 ACAGCACCACCTAGGTACCGCCAGAGCGGGCTGAGCAGGAGGCCCT 348
334 eLysGlyGluGluGlnGluLeuGluLeuAsnThrLysLeuLysThrL 351
   : ||||| ::||| ::||| ::|||
349 CACTGGCGAGTGTAAAGAGAGCTAGAAAAAAGAAATGAGGCTCTGAAAGAGA 398
351 ysValAspLeuGlnArgGluIleLysTyrMetLysAspLeuMetGlu 367
   ||||| ::||| ::||| ::||| ::||| ::||| ::|||
399 AGCAGATTCTTCGCCAAGAGATCCAGTATCTGAAAGACCTGTAGAA 448
368 AspValCysLysAlaLysGly 374
   ::||| ::|||
449 GAGGTCCGTAAAGCAAGGGG 469
```